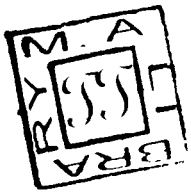

ON USE OF MATHEMATICAL PROGRAMMING TECHNIQUES IN SOME OPTIMIZATION PROBLEMS ARISING IN STRATIFIED SAMPLE SURVEYS

ABSTRACT
THESIS

SUBMITTED FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

STATISTICS

BY

EHSAN AHMAD KHAN

UNDER THE SUPERVISION OF
DR. MOHAMMAD JAMEEL AHSAN

DEPARTMENT OF STATISTICS & OPERATIONS RESEARCH
ALIGARH MUSLIM UNIVERSITY, ALIGARH 202002, INDIA
1998

ABSTRACT

This thesis entitled **"ON USE OF MATHEMATICAL PROGRAMMING TECHNIQUES IN SOME OPTIMIZATION PROBLEMS ARISING IN STRATIFIED SAMPLE SURVEYS"** is submitted to the Aligarh Muslim University, Aligarh, to supplicate the degree of Doctor of Philosophy in Statistics. It embodies the research work carried out by me at the Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh.

There are two generally accepted options for studying the characteristics of a finite population. The first is to study each and every unit of the population. This is called census or complete enumeration. The other is through the study of only a selected portion of the population. This selected portion is called a 'sample' and this method is known as sampling or sample survey. Census is time-consuming, expensive, even impossible in some situations and often inaccurate. On the other hand a sample survey costs less in terms of time and money both and usually is more accurate than the census.

One of the extensively used sampling designs (method of sample survey) is the stratified sampling. In this sampling design the population are divided into non-overlapping exhaustive and as

of required sizes are then selected from each strata. Apart from increasing the precision of the estimates, stratified sampling may also provide estimates for the different subdivisions constituting the population.

In this thesis, some constrained optimization problems arising in univariate and multivariate stratified sample surveys are discussed. These problems are formulated as mathematical programming problems and special purpose algorithms are developed for solving them using mathematical programming techniques.

This thesis consists of FIVE chapters. Chapter-I provides an introduction to the basic ideas in sampling surveys specifically in simple random sampling and stratified sampling designs. An introduction to mathematical programming and its application to solve various problems arising in stratified sample surveys is also presented in this chapter.

Chapter-II deals with the problem of determining the optimum number of strata. The problem is studied under three different situations. These problems are formulated as non-linear mathematical programming problems and their solutions are obtained with the help of the well known Kuhn-Tucker condition of non-linear programming. This chapter is based on my joint research paper Khan et al (1998) appeared in "Frontiers in Probability and Statistics", edited by S.P. Mukherjee, S.K. Basu and B.K. Sinha

and published by Narosa Publishing House, New Delhi.

In Chapter-III the problem of determining the optimum strata boundaries is studied and a new formulation of the problem is provided in the form of a mathematical programming problem. Dynamic programming technique is then used to work out the optimum strata boundaries. Two numerical examples are also presented to illustrate the computational details of the procedure developed for solution. This chapter is based on my joint research paper "Optimum Stratification: A Mathematical Programming Approach", accepted for presentation in the VI Islamic Society of Statistical Science Conference (ISOSSC) to be held in Dhaka during December 12-15, 1998.

The most important problem in stratified sampling is the determination of sample sizes (allocations) for different strata. They may be chosen to minimize the sampling variance of the estimator for a fixed cost or to minimize the total cost of the survey for a desired precision. Such an allocation is called an optimum allocation. In Chapter-IV the above problem is formulated as a mathematical programming problem in a situation where p different characteristics are defined on every population unit. A procedure for solving this problem is developed by using dynamic programming technique. This chapter is based on my joint research paper entitled "On compromise allocation in multivariate

stratified sampling", submitted for publication to Naval Research Logistics (vide their manuscript number 3280).

The fifth and last chapter provides an integer solution of the problem worked out in Chapter-IV. This chapter is based on my joint research paper entitled "An optimal multivariate stratified sampling design using dynamic programming", presented in the "3rd International Triennial Calcutta Symposium" held in December 1997. The paper is due to appear in the proceedings of the symposium to be published by Wiely Publications.

A comprehensive list of references, arranged in alphabetical order is also provided at the end of the thesis.

ON USE OF MATHEMATICAL PROGRAMMING TECHNIQUES IN SOME OPTIMIZATION PROBLEMS ARISING IN STRATIFIED SAMPLE SURVEYS

THESIS

SUBMITTED FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

STATISTICS

BY

EHSAN AHMAD KHAN

UNDER THE SUPERVISION OF

DR. MOHAMMAD JAMEEL AHSAN

DEPARTMENT OF STATISTICS & OPERATIONS RESEARCH
ALIGARH MUSLIM UNIVERSITY, ALIGARH 202002, INDIA

1998

Tel. : 401251

DEPARTMENT OF
STATISTICS & OPERATIONS RESEARCH
ALIGARH MUSLIM UNIVERSITY
ALIGARH—202 002, INDIA

CERTIFICATE

I certify that the material contained in this thesis entitled "**ON USE OF MATHEMATICAL PROGRAMMING TECHNIQUES IN SOME OPTIMIZATION PROBLEMS ARISING IN STRATIFIED SAMPLE SURVEYS**" by Mr.Ehsan Ahmad Khan for the award of the degree of Doctor of Philosophy in Statistics is original.

The work has been done under my supervision. In my opinion the work contained in this thesis is sufficient for consideration for the award of a Ph.D. degree in Statistics.

Dated: 7.11.98

(DR. MOHAMMAD JAMEEL AHSAN)
Supervisor,
Reader,
Department of Statistics and
Operations Research,
Aligarh Muslim University,
Aligarh-2020002, INDIA.

Chairman
Dept. of Stats & Oper. Research
Aligarh Muslim University, Aligarh

PREFACE

This thesis entitled **"ON USE OF MATHEMATICAL PROGRAMMING TECHNIQUES IN SOME OPTIMIZATION PROBLEMS ARISING IN STRATIFIED SAMPLE SURVEYS"** is submitted to the Aligarh Muslim University, Aligarh, to supplicate the degree of Doctor of Philosophy in Statistics. It embodies the research work carried out by me at the Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh.

There are two generally accepted options for studying the characteristics of a finite population. The first is to study each and every unit of the population. This is called census or complete enumeration. The other is through the study of only a selected portion of the population. This selected portion is called a 'sample' and this method is known as sampling or sample survey. Census is time-consuming, expensive, even impossible in some situations and often inaccurate. On the other hand a sample survey costs less in terms of time and money both and usually is more accurate than the census.

One of the extensively used sampling designs (method of sample survey) is the stratified sampling. In this sampling design the population are divided into non-overlapping exhaustive and as

of required sizes are then selected from each strata. Apart from increasing the precision of the estimates, stratified sampling may also provide estimates for the different subdivisions constituting the population.

In this thesis, some constrained optimization problems arising in univariate and multivariate stratified sample surveys are discussed. These problems are formulated as mathematical programming problems and special purpose algorithms are developed for solving them using mathematical programming techniques.

This thesis consists of FIVE chapters. Chapter-I provides an introduction to the basic ideas in sampling surveys specifically in simple random sampling and stratified sampling designs. An introduction to mathematical programming and its application to solve various problems arising in stratified sample surveys is also presented in this chapter.

Chapter-II deals with the problem of determining the optimum number of strata. The problem is studied under three different situations. These problems are formulated as non-linear mathematical programming problems and their solutions are obtained with the help of the well known Kuhn-Tucker condition of non-linear programming. This chapter is based on my joint research paper Khan et al (1998) appeared in "Frontiers in Probability and Statistics", edited by S.P. Mukherjee, S.K. Basu and B.K. Sinha

and published by Narosa Publishing House, New Delhi.

In Chapter-III the problem of determining the optimum strata boundaries is studied and a new formulation of the problem is provided in the form of a mathematical programming problem. Dynamic programming technique is then used to work out the optimum strata boundaries. Two numerical examples are also presented to illustrate the computational details of the procedure developed for solution. This chapter is based on my joint research paper "Optimum Stratification: A Mathematical Programming Approach", accepted for presentation in the VI Islamic Society of Statistical Science Conference (ISOSSC) to be held in Dhaka during December 12-15, 1998.

The most important problem in stratified sampling is the determination of sample sizes (allocations) for different strata. They may be chosen to minimize the sampling variance of the estimator for a fixed cost or to minimize the total cost of the survey for a desired precision. Such an allocation is called an optimum allocation. In Chapter-IV the above problem is formulated as a mathematical programming problem in a situation where p different characteristics are defined on every population unit. A procedure for solving this problem is developed by using dynamic programming technique. This chapter is based on my joint research paper entitled "On compromise allocation in multivariate

stratified sampling", submitted for publication to Naval Research Logistics (vide their manuscript number 3280).

The fifth and last chapter provides an integer solution of the problem worked out in Chapter-IV. This chapter is based on my joint research paper entitled "An optimal multivariate stratified sampling design using dynamic programming", presented in the "3rd International Triennial Calcutta Symposium" held in December 1997. The paper is due to appear in the proceedings of the symposium to be published by Wiely Publications.

A comprehensive list of references, arranged in alphabetical order is also provided at the end of the thesis.

ACKNOWLEDGEMENT

On the eve of completion of this research work first of all I would like to express my indebtedness and sincere gratitude towards my supervisor, Dr. Mohammad Jameel Ahsan, Reader, Department of Statistics, Aligarh Muslim University, Aligarh for his invaluable guidance and constant encouragement through out the course of this research work. His great involvement and sympathetic behaviour was the main cause behind this work.

I am extremely grateful to Prof. Abdul Hameed Khan, Chairman, Department of Statistics and Operations Research who very kindly provided me with necessary facilities to complete this work. I am also grateful to Prof. S.Rehman, Prof. S.U. Khan, Prof. M.Z. Khan for their encouragement which was a constant source of inspiration to me.

On the occasion of submitting this work my sense of gratitude is due to my beloved teacher (late) Prof. Zaheeruddin whose untimely demise, a few weeks before the submission of this thesis, left a void not only in his family and his department but also in the depth of my heart. May his soul rest in peace. Aameen.

I am also thankful to all the members of the department of Statistics and Operations Research, A.M.U., Aligarh for their kind help and co-operation.

I am full of gratitude to Dr. Md. Golam Mostafa Khan, Lecturer in Statistics, Mohd. Sathak, college of Arts & Science, Sholinganallur, Madras for his sincere and continuous support towards completion of this manuscript. Thanks are also due to Dr. Z.A. Khan, Centre of Creative Education, Osaka, Japan and Dr.(Mrs) N. Jahan, Associate Professor, Department of Statistics, Dhaka University, Dhaka, Bangladesh for their kind support and encouragement.

I would like to express my gratefulness to Prof. Alauddin Ahmad, Vice-Chancellor, Hamdard University and Prof. Mohammad Athar, Head, Department of Medical Elementology and Toxicology, Faculty of Science, Hamdard University, NEw Delhi, India for their encouragement and moral support, which enabled me to complete this research work and to write this thesis.

I hereby express my sincere gratitude to all my family members, my parents, my brothers, specially Mr.Kamal Ahmad Khan and Mr. Riaz Ahmad Khan and my wife Mrs. Rabia Khatoon for their constant encouragement, assistance and invocation which enabled me to reach this stage of my life and to complete this work.

My words of thanks are also due to Mr. Israr Ahmad of Jamia Millia Islamia University, New Delhi for typing this manuscript.

E Akhan

Date : 7.11.98

(Ehsan Ahmad Khan)
Department of Statistics
and Operations Research
Aligarh Muslim University
Aligarh-202002, INDIA.

CONTENTS

CHAPTER-1

INTRODUCTION

1.1 SAMPLE SURVEYS

The information about a population can be collected either by conducting a census or a sample survey. A census or complete enumeration is that in which all the elements (units) constituting the population are studied and conclusions are drawn therefrom. On the other hand in a sample survey only a selected portion of the population, called sample is selected and studied and the estimates for population characteristics such as population mean, population total, population proportion, population variance etc., are constructed on the basis of the sample observations. Sampling, that is, the selection of a part of an aggregate of material to represent the whole aggregate is a long established practice. A sampling method is a scientific and objective procedure of selecting units from the population and provides a sample that is expected to be a representative of the population as a whole. A sampling method makes it possible to estimate the population parameters such as population total, average or proportion while the size of survey operations are considerably reduced as compared to census. The aim of a sample survey is the collection of information to satisfy a definite need. The need to collect the

information about population characteristics arises in every conceivable sphere of human activity.

A sample survey is less costly than a complete enumeration because the expenses of observing all units should obviously be greater than that of observing only a small portion. Also it takes less time to collect and process the data from a sample than that of a census data. The results obtained from a carefully planned and well executed sample survey are expected to be more accurate than those of a complete census. A complete census ordinarily requires a huge and unwieldy organization and therefore many types of errors creep in, which cannot be controlled adequately. In a sample survey the volume of work is reduced considerably and it becomes possible to employ persons of high caliber, train them suitably and supervise their work effectively. In sample surveys it is possible to make a valid estimate of the margin of error, and hence to decide the accuracy of the result. Thus sampling enquiries are becoming more and more popular in all spheres of human activity. They are specially advantageous in case of social surveys. The large universe (population), difficulties in contacting people, high non-response rate etc., makes sampling the best procedure in case of social investigations. Recent developments in the science of statistics, specially in the field of sampling, have made these procedures more realistic and reliable. In the planning of sample surveys the sample involves fewer respondents than a census, for which all units in the field

covered are respondents. Practically no one has time or means to make a complete investigation for every problem with which he comes into contact. He must therefore rely on sampling. The aim of sampling methods is to obtain maximum information about the phenomenon under study with minimum possible use of money, time and energy.

1.2 PROBABILITY SAMPLING

A sampling procedure which satisfies the following properties is termed as 'Random' or 'Probability Sampling'.

(i) A set of distinct samples $S_1, S_2, \ldots, S_k$ of a fixed size can be defined.

(ii) Each possible sample S_i is assigned a known probability of selection p_i, i=1,2,,...,k.

(iii) The sampling procedure is capable of selecting any one of the possible sample S_i with its assigned probability p_i.

(iv) The method of computing the estimate from the sample leads to a unique estimate for any specified sample.

A sampling procedure which does not satisfy the above properties is termed as non-probability sampling.

Since no elements of probability is involved in non-probability sampling procedures, they are not capable of further development of the theory. Hence in this manuscript, hereinafter, by sampling we mean random sampling.

1.3 SIMPLE RANDOM SAMPLING (SRS)

It is the simplest form of the random sampling in which all possible samples of a given size have equal chance of being selected.

If a simple random sample of size n is to be drawn from a population of size N there will be

$$ {}^{N}C_{n} = \frac{N!}{n!(N-n)!} \quad \text{possible samples.} $$

Let y_i = the measurement on the i^{th} unit of the population/sample.

$$ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{(the sample mean)} $$

and

$$ \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i \quad \text{(the population mean)}. $$

It is well known that the sample mean $\bar{y}$ is an unbiased estimate of the population mean $\bar{Y}$ with a sampling variance.

$$ V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2 $$

where $S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$ is the population variance of y_i.

An unbiased estimate of S^2 is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Thus $v(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)s^2$ provides us an unbiased estimate of $V(\bar{y})$.

1.4 STRATIFIED SAMPLING

In stratified sampling the population is first divided into groups called strata. These strata are mutually exclusive and exhaustive. Independent simple random samples are then drawn from these strata.

The procedure of stratified sampling is intended to give a better cross-section of the population than that of unstratified sampling. It follows that one would expect the precision of the estimates of the population characteristics to be higher in stratified than in unstratified sampling. Stratified sampling is also convenient in other ways like the selection of sampling units, the location and enumeration of the selected units, distribution and supervision of field-work. In general the whole administration of the survey is greatly simplified in stratified sampling.

Let the population of size N be divided into L

strata of sizes $N_1, N_2, \ldots, N_L$ such that the strata are mutually exclusive and $\sum_{h=1}^{L} N_h = N$. Furthermore let simple random samples of sizes $n_1, n_2, \ldots, n_L$ have been drawn independently from $1^{st}, 2^{nd}, \ldots, L^{th}$ stratum respectively.

Let the measurement on the j^{th} unit of the h^{th} stratum be y_{hj}.

For the h^{th} stratum

N_h = Stratum size

n_h = Sample size

$W_h = \frac{N_h}{N}$ = Stratum weight or stratum proportion

$f_h = \frac{n_h}{N_h}$ = Sampling fraction

$\bar{Y}_h = \frac{1}{N_h} \sum_{j=1}^{N_h} y_{hj}$ = Stratum mean

$\bar{y}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj}$ = Sample mean

$S_h^2 = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (y_{hj} - \bar{Y}_h)^2$ = Stratum variance

$s_h^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2$ = Sample variance.

Also

$$\bar{Y} = \frac{1}{N} \sum_{h=1}^{L} \sum_{j=1}^{N_h} Y_{hj} = \sum_{h=1}^{L} W_h \bar{Y}_h = \text{Overall population mean.}$$

It is well known that the stratified sample mean

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{1.1}$$

is an unbiased estimate of $\bar{Y}$ with a sampling variance

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2. \tag{1.2}$$

There are many more sampling designs e.g. Cluster sampling, Systematic sampling, Multistage sampling etc., but these are beyond the scope of this manuscript. In this manuscript the optimization problems arising in stratified sampling and their possible solutions using mathematical programming techniques are studied.

1.5 AUXILIARY INFORMATION IN SAMPLE SURVEYS

Any variable which is highly correlated with the main estimation variable and whose measurements are either available or relatively cheaper can be used to increase the precision of the estimates such a variable is termed as an auxiliary variable.

The auxiliary information may be used in many ways, like construction of ratio and regression estimates and in assigning the probabilities of inclusion in the sample to various units of a population. Stratification can also be done with the help of an auxiliary variable.

1.6 OPTIMIZATION PROBLEMS ARISING IN STRATIFIED SAMPLING

The practical implementation of stratified sampling in any sample survey requires the solution of the following three basic optimization problems:

(a) the determination of the optimum number of strata
(b) the determination of the optimum strata boundaries
(c) the determination of the optimum allocations to various strata.

The solutions of all the above problems for univariate case i.e. when a single characteristics is studied on each and every population unit, exist in sampling literature. However the multivariate case is more complicated and few attempts have been made to attack the above problems so far.

In the Chapter-II and Chapter-III of this thesis the first two problems are formulated as a mathematical programming problem (MPP) in univariate cases and specialized MPP techniques are developed to solve them.

In the Chapter-IV and Chapter-V of this thesis the problem of

optimum allocation in multivariate stratified sampling is formulated as a mathematical programming problem and solution procedures are developed using dynamic programming techniques.

1.7 MATHEMATICAL PROGRAMMING

Any problem seeking maximization or minimization of a function of one or more variables where the variables are independent or related in some way may be referred to as an optimization problem. Since the past 200 years differential calculus and calculus of variations are in use to solve certain type of optimization problems.

In the last 50 years a new class of optimization problems are emerged. These problems are the real life optimization problems that are usually not amenable to solution by the classical method of calculus. These problems are termed as Mathematical Programming Problems (MPP) by Robert Dorfman around 1950. A wide variety of optimization problems arising in engineering, operations research, management science, economics, military operations, industry, agriculture, statistics etc. are formulated as MPP and specific algorithms are developed to solve them.

The mathematical model of the general MPP may be given as follows:

Minimize $f(\underline{x})$ (1.3)

Subject to $g_i(\underline{x}) \leq \text{or} = \text{or} \geq b_i$, $i=1,2,\ldots,m.$ (1.4)

where $\underline{x}' = (x_1, x_2, \ldots, x_n)$ is the vector of unknown decision variables and f and $g_i (i=1,2,\ldots,m)$ are real valued functions of the n real decision variables $x_1, x_2, \ldots, x_n$.

The function $f(\underline{x})$ is called the objective function and the set of m restrictions in (1.4) are referred to as constraints. Further more one and only one of the signs $\leq$, $=$, $\geq$ holds for each constraint, but the sign may vary from one constraint to other.

Usually all the decision variables are restricted to be non-negative. If this is the case we may add the non-negativity restrictions

$$x_j \geq 0; \quad j=1,2,\ldots,n \tag{1.5}$$

to the above formulation.

In the above formulation the MPP is stated as a minimization problem. This could be done without any loss of generality because of the fact that maximization of $f(\underline{x})$ is equivalent to minimization of $(-f(\underline{x}))$ and we have

$$\text{maximum of } f(\underline{x}) = - \text{ minimum of } (-f(\underline{x})).$$

The functions $f(\underline{x})$ and $g_i(\underline{x})$ are usually assumed to be continuously differentiable to make the problem more tractable to theoretical treatment.

If all the function appearing in an MPP are linear functions

of decision variables x_j; $j=1,2,\ldots,n$ the MPP is called a linear programming problem (LPP) otherwise it is called a non-linear programming problem (NLPP). Further more if all the decision variables are restricted to be integers the LPP or NLPP, as the case may be, is called an all integer LPP (AILPP) or an all integer NLPP (AINLPP) accordingly.

The following form of the MPP may be taken as its standard form

$$\text{Minimize} \qquad f(\underline{x}) \tag{1.6}$$

$$\text{Subject to} \qquad g_i(\underline{x}) \geq 0; \quad i=1,2,\ldots,n \tag{1.7}$$

$$\text{and} \qquad \underline{x} \geq \underline{0}. \tag{1.8}$$

Any $\underline{x}'=(x_1,x_2,\ldots,x_n)$ satisfying (1.7) and (1.8) is called a feasible solution to MPP (1.6)-(1.8). The collection of all feasible solutions of an MPP is called its set of feasible solutions. Usually the set of feasible solutions is denoted by F. Thus:

$$F = \{\underline{x} \mid g_i(\underline{x}) \geq 0;\ i=1,2,\ldots,m \text{ and } x_j \geq 0;\ j=1,2,\ldots,n\}$$

Any $\underline{x}^* \in F$ is called an optimal solution to the MPP (1.6)-(1.8) if $f(\underline{x}^*) \leq f(\underline{x})$ for all $\underline{x} \in F$.

1.8 COMPUTATIONAL PROCEDURES FOR SOLVING MATHEMATICAL PROGRAMMING PROBLEMS

The problems of planning and co-ordination among various project and optimum allocation of limited resources to obtain the desired result were emerged as the basic problems during and after World War II. Intensive work by the United States Air Force team SCOOP (Scientific Computation of Optimum Programs) led by George B.Dantzig resulted in the development of the famous Simplex Algorithm for solving LPP. Simplex method is an iterative procedure which yields an exact optimal solution in a finite number of steps.

By suitable transformations some non-linear programming problems may also be converted into a form which permits the use of simplex algorithm. Thus simplex algorithm emerges as the most powerful computational device for solving linear as well as some non-linear programming problems.

Kuhn H.W. and Tucker A.W. (1951) derived the necessary conditions (popularly known as the K-T conditions) to be satisfied by an optimal solution to an MPP. These conditions laid the foundation of a great deal of further development of the non-linear programming algorithms.

Like simplex algorithm for solving LPP, till date no single algorithm is available for solving the general NLPP. However

special algorithms are developed for solving NLPP's having certain special features. A brief description of some of these are presented here.

Beal (1959) and Wolfe (1959) developed method for solving Quadratic Programming Problem (QPP). Other methods for solving a QPP are due to Van de Panne and Whinston (1964a, 1964b, 1966), Lemke (1962), Graves (1967), Fletcher (1971), Aggarwal (1974a,1974b), Finkbeiner and Kall (1978), Arshad, Khan and Ahsan (1981), Khan, Ahsan and Khan (1983), Todd (1985), Fukushima (1986), Powell and Yuan (1986), Ben-Daya and Shetty (1990), Kalantari and Bagchi (1990), Yuan (1991), Wei (1992), Benzi (1993), Fletcher (1993), Bomze and Danninger (1993,1994), Anstreicher, Den Hertog and Terlaky (1994). Rosen (1960,1961), Kelly (1960), Goldfarb (1969), Du Wu and Zhang (1990), Lai, Gao and He (1993) developed Gradient methods for solving NLPP's with some special features. These methods are based on the fact that if we move in the direction of the negative gradient of the objective function the rate of decrease in the value of the objective function is maximum.

Another useful computational technique for solving MPP that have some special special features is the Dynamic Programming technique. In this thesis dynamic programming technique is used as the main tool for solving the optimization problems arising in stratified sampling indicated in Section 1.6 of Chapter-I. The

next section deals with the basic ideas of dynamic programming technique.

1.9 THE DYNAMIC PROGRAMMING TECHNIQUE

Dynamic programming is not any special type of mathematical programming problem like quadratic programming or linear programming etc. By dynamic programming we mean the computational algorithm to solve mathematical programming problems that have some special features. The basic principle which led to the dynamic programming technique was enunciated by Richard Bellman (1957). This principle says "An optimal policy has the property that whatever the initial state and initial decisions are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision".

Bellman and Dreyfus (1962), Wachs (1989), Li (1990), Li and Haimes (1990), Wang (1990a, 1990b), Wang and Xing (1990), Chen, Hearn and Lee (1994), Odanaka (1994) and several other made significant contributions to the development and applications of the dynamic programming technique.

In the next section the general nature of the MPP to which dynamic programming technique can be applied is discussed.

1.10 THE GENERAL NATURE OF THE PROBLEM

The problems requiring sequential decision making at

different stages may be called multistage decision problems. The problem of making a set of optimal decisions may be formulated as an MPP. The dynamic programming technique is a procedure which can handle the problem of optimal decision making at various stages of a multistage decision problem. The general nature of the MPPs that can be attacked by this technique may be described as follows.

(i) The MPP can be treated as a multistage decision problem. At each stage the value(s) of one or more decision variables are to be determined.

(ii) The MPP must have the same structure at every stage irrespective of the number of stages.

(iii) At every stage the values of the decision variables and the objective function must depend on a specified set of parameters describing the state of system. These parameters are called the state parameters.

(iv) Same set of state parameters must describe the state of the system irrespective of the number of stages.

(v) The decision at any stage must have no effect on the decisions to be made at the remaining stages except in changing the values of the state parameters.

In solving an MPP by dynamic programming technique we start with a one stage problem, moving on to a two stage problem, to a three stage problem and so on until all stages are included. The final solution is obtained by adding the n^{th} (final) stage to the

solution of (n-1) stage. For this a relation between the two successive stages is defined. This relation is called the "Recurrence relation of dynamic programming".

1.11 THE PROBLEM OF DIMENSIONALITY

The computational efficiency of the dynamic programming technique as compared to the complete enumeration is very impressive. Hadley (1964) showed by means of an artificial example involving 5 variables and only one equality constraints where the variable may assume any of the 21 possible values, that the complete enumerational involves the evaluation of the objective function at 10,626 different five dimensional points as compare to 945 points while using the dynamic programming technique. The computational efforts involved in solving an MPP by dynamic programming technique multiply incredibly fast with the increase in the number of state parameters (number of constraints). The number of state parameters is called the dimensionality of the MPP Thus the problem of handling the great bulk of computation is termed as the "Problem of Dimensionality" or "Curse of Dimensionality".

Bellman and Dreyfus (1962) suggested a procedure to reduce the dimensionality of the problem.

However as far as the problems discussed in this manuscript are concerned dimensionality poses no threat to the convergence of

computational procedures developed using dynamic programming technique.

1.12 USES OF MATHEMATICAL PROGRAMMING

Mathematical programming is widely used in solving a number of problems arising in military, economics, industry, business management, engineering, inventory etc. etc. Mathematical programming is also used, although not as extensively as in above indicated areas, to deal with the problem arising in Statistical Analysis. As pointed out by the C.R. Rao, in Arthanari and Dodge (1981):

"All statistical procedures are, in the ultimate analysis, solutions to suitably formulated optimization problems. Whether it is designing a scientific experiment or planning a large scale survey for collection of data, or choosing a stochastic model to characterize observed data, or drawing inference from available data, such as estimation, testing of hypothesis and decision making, one has to choose an objective function and minimize or maximize it subject to given constraints on unknown parameters and inputs such as the cost involved. The classical optimization methods based on differential calculus are to restrictive and are either inapplicable or difficult to apply in many situations that arise in statistical work. This together with the lack of suitable numerical algorithms for solving optimizing equations has placed several limitations on the choice of objective functions and

constraints and led to the development and use of some inefficient statistical procedures.

Attempts have therefore been made during the last three decades to find other optimization techniques that have wider applicability and can be easily implemented with the available computing power. One such technique that has the potential for increasing the scope for application of efficient statistical methodology is mathematical programming. Although endowed with a vast literature, this method has not come into regular use in statistical practice mainly because of lack of good expositions integrating the techniques of mathematical programming with statistical concepts and procedures".

1.13 MATHEMATICAL PROGRAMMING IN SAMPLING

As stated earlier sampling theory deals with the problems associated with the selection of samples from a population according to certain probability mechanism. A sampling design is a probability measure on the set of all possible samples of a fixed size from a given population. The problem of deriving statistical information about population parameters on the basis of a sample data in the form of estimates can be formulated as a mathematical programming problem. The objective of these problems may be to minimize the cost of the survey as a function of the sample size, the size of the sampling unit, the sampling scheme, the scope of the survey etc. subject to certain limits on the loss in precision

due to the decisions made on the basis of sample data. One may also be interested in maximizing the precision of the estimate based on a sample, subject to cost restrictions.

In this thesis an attempt is being made to use mathematical programming technique to solve some problems arising in stratified sampling.

CHAPTER-II

DETERMINATION OF THE OPTIMUM NUMBER OF STRATA

2.1 INTRODUCTION

In stratified sampling it seems that if the population of a given size is divided into a large number of strata of small sizes then the efficiency due to the stratification also increases. Analysis of the results shows that this is not true. For a given value of total sample size n there is a value of L (the optimum number of strata) beyond which any further increase in the number of strata may not improve the efficiency of stratification. Thus the value of determining the optimum number of strata assumes considerable importance.

In this chapter the problem of determining the optimum number of strata is discussed as a non-linear programming problem. Three different objective functions are considered which resulted into three separate nonlinear programming problems. The solution of these problems are worked out either by using the classical lagrange multipliers technique or by using the well known Kuhn and Tucker necessary conditions of mathematical programming problem.

This chapter is based on my joint research paper Khan et al (1998) appeared in "Frontiers in Probability and Statistics"

edited by S.P. Mukherjee, S.K. Basu and B.K. Sinha and published by "Narosa Publishing House, New Delhi".

2.2 THE KUHN AND TUCKER CONDITIONS

Kuhn H.W. and Tucker A.W. (1951) gave a set of necessary conditions to be satisfied by an optimal solution of a non-linear programming problem.

Consider the NLPP in the form

$$\begin{aligned} &\text{Maximize} && f(\underline{x}), \\ &\text{Subject to} && g_i(\underline{x}) \geq 0; \quad i=1,2,\ldots,m \qquad (2.1) \\ &\text{and} && \underline{x} \geq \underline{0} \end{aligned}$$

where f and g_i; $i=1,2,\ldots,m$ are functions of n decision variables $x_1, x_2, \ldots, x_n$ and $\underline{x}' = (x_1, x_2, \ldots, x_n)$ denote the n component vector of decision variables.

Define the function ϕ as

$$\phi(\underline{x},\underline{u}) = f(\underline{x}) + \sum_{i=1}^{m} u_i g_i(\underline{x})$$

where $\underline{u}' = (u_1, u_2, \ldots, u_m)$ is an m-component vector of multipliers.

Let $\nabla_{\underline{x}} \phi(\underline{x},\underline{u})$ and $\nabla_{\underline{u}} \phi(\underline{x},\underline{u})$ denote the gradient vectors i.e. the vectors of partial derivatives of ϕ with respect to the components of $\underline{x}$ and $\underline{u}$ respectively. If $\underline{x}^*$ is an optimal solution

to the NLPP (2.1) then Kuhn and Tucker (1951) showed that under certain regularity conditions there exists a vector $\underline{u}^*$ such that the following conditions hold.

$$\left.\begin{aligned} \nabla_{\underline{x}}\phi(\underline{x}^*,\underline{u}^*) &\le \underline{0}, \\ \underline{x}^{*\prime}\nabla_{\underline{x}}\phi(\underline{x}^*,\underline{u}^*) &= 0, \\ \underline{x}^* &\ge \underline{0}, \\ \nabla_{\underline{u}}\phi(\underline{x}^*,\underline{u}^*) &\ge \underline{0}, \\ \underline{u}^{*\prime}\nabla_{\underline{u}}\phi(\underline{x}^*,\underline{u}^*) &= 0, \\ \underline{u}^* &\ge \underline{0}. \end{aligned}\right\} \qquad (2.2)$$

When the objective function f is pseudoconcave and the constraint functions g_i, i=1,2,...,m are quasiconcave the conditions in (2.2) are sufficient also. Thus if the conditions are necessary and sufficient both and we are able to find $\underline{x}^*$ and $\underline{u}^*$ satisfying the conditions (2.2) then $\underline{x}^*$ will solve the NLPP (2.1).

2.3 FORMULATION OF THE PROBLEMS

The problem of determining the optimum number of strata was first discussed by Dalenius (1950). He used the main variable y as the stratification variable. Dalenius (1953) conjectured that the ratio of the variance $V(\bar{y}_{st})_L$ to the variance $V(\bar{y}_{st})_{L-1}$ of the stratified sample means based on L and L-1 strata respectively is given by $V(\bar{y}_{st})_L \;/\; V(\bar{y}_{st})_{L-1} = \left(\frac{L-1}{L}\right)^2$. Later on Cochran (1961)

confirmed that the above relationship holds approximately for skewed distribution also and apparently the rate of reduction in the variance is independent of the skewness of the population. Later on Sethi (1963) showed that in case of gamma distributions the variance ratio $V(\bar{y}_{st})_L$ / $V(\bar{y})_{SRS}$ can be expressed as the inverse of a quadratic function in L.

In the following the problem of determining the optimum number of strata is formulated as a NLPP in three different situations.

Problem 1: Under certain assumptions Dalenius (1950) postulated that in stratified sampling with L strata the variance $V(\bar{y}_{st})$ of the stratified sample mean $\bar{y}_{st}$ is inversely proportional to L^2, that is

$$V(\bar{y}_{st}) \propto \frac{1}{L^2} \text{ or } V(\bar{y}_{st}) = \frac{A}{L^2}$$

where A is the constant of proportionality. The constant A is seen approximately equal to $\frac{S_y^2}{n}$, where S_y^2 is the population variance of y and n is the total sample size. Thus

$$V(\bar{y}_{st}) \cong \frac{S_y^2}{nL^2} . \qquad (2.3)$$

The cost C involved in making L strata and selecting and evaluating a stratified sample of size n may be given as

$$C = c_1 L + c_2 n \qquad (2.4)$$

where c_1=cost per stratum and c_2=cost per unit within each stratum.

To obtain an unbiased estimate of the stratum variances S_h^2; h=1,2,...,L there should be random samples of at least 2 units from each stratum which is possible if and only if L≤n/2 or

$$2L \leq n. \tag{2.5}$$

The problem is to find L and n (say L^* and n^*) that minimizes (2.3) subject to cost constraint and constraint (2.5). Thus we have the following MPP to solve:

$$\left.\begin{array}{ll} \text{Minimize} & V(n,L) = \dfrac{S_y^2}{nL^2} \\ \text{Subject to} & c_1L+c_2n \leq C_o \\ & 2L \leq n \\ & L \geq 0,\ n \geq 0 \end{array}\right\} \qquad \text{(MPP-I)}$$

where C_o is the total amount available for the survey and the non-negativity restriction L≥0, and n≥0 are introduced because negative values of L and n are of no practical use.

Problem 2: When an auxiliary variable x is used as stratification variable Cochran (1963) showed that if the regression of y (the main variable) on x is linear then for a given value of L

$$V(\bar{y}_{st}) \geq \frac{S_y^2}{n}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] \qquad (2.6)$$

where ρ is the correlation coefficient between x and y in the unstratified population.

In this situation one may be interested in finding L (and n) for which the right hand side (RHS) of (2.6) is minimum, subject to the restrictions discussed in Problem 1 above. Thus we have to solve the following MPP:

$$\left.\begin{array}{ll} \text{Minimize} & V(n,L) = \dfrac{S_y^2}{n}\left[\dfrac{\rho^2}{L^2} + (1-\rho^2)\right] \\ \text{Subject to} & c_1L+c_2n \leq C_o \\ & 2L \leq n \\ & L \geq 0,\ n \geq 0. \end{array}\right\} \qquad \text{(MPP-II)}$$

Problem 3: Sethi (1963) showed that when the main variable y follows a Gamma distribution the variance $V(\bar{y}_{st})$ of the stratified sample mean $\bar{y}_{st}$ may be expressed as

$$V(\bar{y}_{st}) = \frac{S_y^2}{n[aL^2+bL+c]} \qquad (2.7)$$

for proportional and equal allocations.

Where a,b and c are constants to be determined by considering the values of the variance ratio $\left(\frac{S_y^2}{n}\right) / V(\bar{y}_{st})$, obtained for L=1,2,

and 3. The problem of finding the optimum number of strata in this situation may be given as

$$\left.\begin{array}{ll} \text{Minimize} & V(n,L) = \dfrac{S_y^2}{n[aL^2+bL+c]} \\ \text{Subject to} & c_1L+c_2n \le C_o \\ & 2L \le n \\ & L \ge 0,\ n \ge 0. \end{array}\right\} \qquad \text{(MPP-III)}$$

2.4 THE SOLUTIONS

Problem 1 (MPP-I): The set F of feasible solutions of MPP-I is defined as

$$F = \{(n,L) \mid c_1n+c_2L \le C_o,\ 2L \le n,\ L \ge 0,\ n \ge 0\}. \qquad (2.8)$$

Obviously the objective function V(n,L) of MPP-I will assume its minimum when the function $Z(n,L)=nL^2$ will assume its maximum under the given constraints. We will first prove the following result.

Theorem: The optimal solution (n^*,L^*) of MPP-I is a boundary point of F.

Proof: Let $(n^*,L^*) \in F$ be an optimal solution to the MPP-I, that is

$$V(n^*,L^*) \le V(n,L) \text{ for all } (n,L) \in F$$

or

$$Z(n^*,L^*) \ge Z(n,L) \text{ for all } (n,L) \in F. \qquad (2.9)$$

If (n^*,L^*) is not a boundary point we can always find another point $(n^*+\partial n, L^*+\partial L) \in F$ for ∂n and $\partial L>0$, however small.

At the point $(n^*+\partial n, L^*+\partial L)$

$$\begin{aligned} Z(n^*+\partial n, L^*+\partial L) &= (n^*+\partial n)(L^*+\partial L)^2 \\ &= (n^*+\partial n)(L^{*2}+\partial L^2+2L^*\partial L) \\ &= n^*L^{*2}+\text{A positive quantity} \\ &= Z(n^*,L^*)+\text{ A positive quantity} \end{aligned}$$

$$\Rightarrow \quad Z(n^*+\partial n, L^*+\partial L) > Z(n^*,L^*).$$

This contradicts (2.9). Hence (n^*,L^*) must be a boundary point of F.

Corollary: The optimal point (n^*,L^*) must be a point on the boundary of F defined by $c_1L+c_2n=C_o$ or $2L=n$ or on their intersection.

Proof: Obviously if $n=0$ there will be no sampling. Also $L=0$ has no meaning because there must be at least one stratum.

Thus we conclude that (n^*,L^*) will be a point on the boundary $c_1L+c_2n=C_o$ or on the boundary $2L=n$ or on the intersection of the two.

The above results suggests that to solve the MPP-I it is sufficient to solve the MPP:

$$\text{Minimize} \qquad V(n,L) = \frac{S_y^2}{nL^2}$$
$$\text{Subject to} \qquad c_1 n + c_2 L = C_o . \tag{2.10}$$

If the solution of the problem (2.10) satisfies the constraint $2L \leq n$, and the restrictions $L \geq 0$ and $n \geq 0$ then the MPP-I is solved completely.

For solving (2.10) we may use the well known lagrange multipliers technique of constrained optimization. The lagrangian function $\psi(n,L,\lambda)$, where λ is the lagrange multiplier, for problem (2.10) is defined as

$$\psi(n,L,\lambda) = \frac{S_y^2}{nL^2} + \lambda(c_1 L + c_2 n - C_o) . \tag{2.11}$$

Differentiating $\psi(n,L,\lambda)$ with respect to n,L and λ partially and equating to zero we get the following three equations

$$\frac{\partial \psi}{\partial n} = - \frac{S_y^2}{n^2 L^2} + c_2 \lambda = 0 \tag{2.12}$$

$$\frac{\partial \psi}{\partial L} = - \frac{2S_y^2}{nL^3} + c_1 \lambda = 0 \tag{2.13}$$

$$\frac{\partial \psi}{\partial \lambda} = c_1 L + c_2 n - C_o = 0 \tag{2.14}$$

(2.12) and (2.13) together give

$$L = \frac{2c_2 n}{c_1} . \qquad (2.15)$$

Substituting this value of L in (2.14) we get

$$c_1\left(\frac{2c_2 n}{c_1}\right)+c_2 n-C_o = 0$$

or

$$n = \frac{C_o}{3c_2} . \qquad (2.16)$$

Substituting this value of n in (2.15) we get

$$L = \left(\frac{2c_2}{c_1}\right)\left(\frac{C_o}{3c_2}\right)$$

or

$$L = \frac{2C_o}{3c_1} . \qquad (2.17)$$

As stated earlier, if the values of n and L given by (2.16) and (2.17) respectively, satisfies the remaining constraint $2L \leq n$ and the non-negativity restriction $L \geq 0$ and $n \geq 0$ the MPP-I is solved and the optimal solution is given

$$L^* = \frac{2C_o}{3c_1} \qquad (2.18)$$

and

$$n^* = \frac{C_o}{3c_2} . \qquad (2.19)$$

Now (2.16) and (2.17) will satisfy $2L \leq n$ if and only if

$$\frac{4C_o}{3c_1} \leq \frac{C_o}{3c_2}$$

or

$$\frac{c_1}{c_2} \geq 4. \tag{2.20}$$

Therefore if the cost ratio $\frac{c_1}{c_2}$ satisfies (2.20), MPP-I is completely solved and the optimal values of L and n that is L^* and n^* are given by (2.18) and (2.19) respectively.

Now if

$$\frac{c_1}{c_2} < 4 \tag{2.21}$$

the constraint $2L \leq n$ is violated. In this situation at the optimal point both the constraints $c_1L+c_2n \leq C_o$ and $2L \leq n$ will be active. There is only one such point that is the point of intersection of the straight lines $c_1L+c_2n=C_o$ and $2L=n$.

Solving

$$c_1L+c_2n = C_o$$

and

$$2L = n$$

as simultaneous equations we get the optimal values of L and n as

$$L^* = \frac{C_o}{c_1+2c_2} \tag{2.22}$$

and

$$n^* = \frac{2C_o}{c_1+2c_2} \quad . \tag{2.23}$$

The following discussion shows that the solutions given by (2.18) and (2.19) when $\frac{c_1}{c_2} \geq 4$ and given by (2.22) and (2.23) when $\frac{c_1}{c_2} < 4$ will provide the required global minimum.

The Hessian matrix (matrix of second order partial derivatives) of the objective function $V(n,L) = \frac{S_y^2}{nL^2}$ is

$$H_V = \begin{bmatrix} \frac{\partial^2 V}{\partial n^2} & \frac{\partial^2 V}{\partial n \partial L} \\ \frac{\partial^2 V}{\partial L \partial n} & \frac{\partial^2 V}{\partial L^2} \end{bmatrix} = \begin{bmatrix} \frac{2S_y^2}{n^3L^2} & \frac{2S_y^2}{n^2L^3} \\ \frac{2S_y^2}{n^2L^3} & \frac{6S_y^2}{nL^4} \end{bmatrix}$$

As both the principal minors of H_V viz.

$$|H_{11}| = \frac{2S_y^2}{n^3L^2}$$

and

$$|H_{22}| = |H_V| = \frac{8S_y^2}{n^4L^6}$$

are greater than zero the function $V(n,L) = \frac{S_y^2}{nL^2}$ is strictly convex for positive values of n. The constraints $c_1L+c_2n=C_o$ and $2L \leq n$ are linear. Thus the values of L and n given by (2.18), (2.19) and

(2.22), (2.23) will provide a global minimum of the MPP-I in the two situations discussed earlier.

It can also be verified that the K-T conditions, which are sufficient also for MPP-I are satisfied by the optimal solutions (2.18), (2.19) and (2.22), (2.23) to MPP-I in situations when $\frac{c_1}{c_2} \geq 4$ and $\frac{c_1}{c_2} < 4$ respectively.

Problem 2 (MPP-II): Consider MPP-II in the following form.

$$\left.\begin{array}{ll} \text{Maximize} & V'(n,L) = -\frac{S_y^2}{n}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] \\ \text{Subject to} & C_o - c_1 L - c_2 n \geq 0 \\ & n - 2L \geq 0 \\ & L \geq 0, \quad n \geq 0. \end{array}\right\} \qquad (2.24)$$

Note that the objective is changed into minimization and the constraints are expressed in "≥0" form.

The Hessian matrix of the objective function $V'(n,L)$ is

$$H_{V'} = \begin{bmatrix} \frac{\partial^2 V'}{\partial n^2} & \frac{\partial^2 V'}{\partial n \partial L} \\ \frac{\partial^2 V'}{\partial L \partial n} & \frac{\partial^2 V'}{\partial L^2} \end{bmatrix} = \begin{bmatrix} -\left(\frac{2S_y^2\rho^2}{n^3L^2} - \frac{2S_y^2\rho^2(1-\rho^2)}{n^3}\right) & -\left(\frac{2S_y^2\rho^2}{n^2L^3}\right) \\ -\left(\frac{2S_y^2\rho^2}{n^2L^3}\right) & -\left(\frac{6S_y^2\rho^2}{nL^4}\right) \end{bmatrix}.$$

The two principal minors of $H_{V'}$ are

$$|H_{11}| = -\frac{2S_y^2\rho^2}{n^3L^2} - \frac{2S_y^2\rho^2(1-\rho^2)}{n^3} < 0$$

and

$$|H_{22}| = |H_{V'}| = \frac{12S_y^4\rho^4}{n^4L^6} + \frac{12S_y^4\rho^2(1-\rho^2)}{n^4L^4} - \frac{4S_y^4\rho^4}{n^4L^6}$$

$$= \frac{8S_y^4\rho^4}{n^4L^6} + \frac{12S_y^4\rho^2(1-\rho^2)}{n^4L^4} > 0.$$

As $|H_{11}|<0$ and $|H_{22}|>0$ the function $V'(n,L)$ is strictly concave. The objective function $V'(n,L)$ of MPP (2.24) is concave and the constraints are linear, hence K-T necessary conditions stated in section 2.2 are sufficient also.

As before first consider the problem for (2.24)

$$\left.\begin{array}{ll} \text{Maximize} & V'(n,L) = -\frac{S_y^2}{n}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] \\ \text{Subject to} & C_o - c_1L - c_2n \geq 0 \\ & L \geq 0, \quad n \geq 0. \end{array}\right\} \qquad (2.25)$$

If the optimal solution of (2.25) satisfies the constraint $n \geq 2L$ of (2.24), MPP-II is solved, otherwise we have to consider the constraint $n \geq 2L$ also.

As the K-T conditions are sufficient also for MPP (2.25), if

we can find $\underline{x}^*$ and $\underline{u}^*$ satisfying (2.2) then $\underline{x}^*$ will solve MPP (2.25).

· The conditions (2.2) for MPP (2.25) are

$$\left.\begin{aligned}
&\nabla_{(n,L)}\theta = \begin{bmatrix} \frac{S_y^2}{n^2}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] - uc_2 \\ \frac{2S_y^2\rho^2}{nL^3} - uc_1 \end{bmatrix} \geq \underline{0} \quad \text{(a)} \\
&(n,L)\nabla_{(n,L)}\theta = n\left[\frac{S_y^2}{n^2}\left\{\frac{\rho^2}{L^2} + (1-\rho^2)\right\} - uc_2\right] \\
&\qquad + L\left[\frac{2S_y^2\rho^2}{nL^3} - uc_1\right] = 0 \quad \text{(b)} \\
&\qquad \begin{pmatrix} n \\ L \end{pmatrix} \geq \underline{0} \quad \text{(c)} \\
&\qquad \nabla_u\theta = (C_o - c_1L - c_2n) \geq 0 \quad \text{(d)} \\
&\qquad u\nabla_u\theta = u(C_o - c_1L - c_2n) = 0 \quad \text{(e)} \\
&\text{and} \qquad u \geq 0 \quad \text{(f)}
\end{aligned}\right\} \quad (2.26)$$

where $\theta(n,L,u) = -\frac{S_y^2}{n}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] + u(C_o - c_1L - c_2n)$

As n and L can not be zero (2.26(b)) implies that

$$\frac{S_y^2}{n^2}\left\{\frac{\rho^2}{L^2} + (1-\rho^2)\right\} - uc_2 = 0$$

or $$u = \frac{S_y^2}{c_2 n^2}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] > 0 \qquad (2.27)$$

and $$\frac{2S_y^2\rho^2}{nL^3} - uc_1 = 0$$

or $$u = \frac{2S_y^2\rho}{c_1 nL^3} > 0 \qquad (2.28)$$

Thus u given by (2.27) and (2.28) satisfies K-T condition (2.26(f)). As u>0 (2.26(e)) implies that $C_o - c_1 L - c_2 n = 0$

or $$n = \frac{C_o - c_1 L}{c_2} \qquad (2.29)$$

Again from (2.27) and (2.28) we get

$$\frac{S_y^2}{c_2 n^2}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] = \frac{2S_y^2\rho}{c_1 nL^3} \qquad (2.30)$$

Equations (2.29) and (2.30) together on simplification gives

$$L^3 + 3\left(\frac{\rho^2}{1-\rho^2}\right)L - 2\left(\frac{\rho^2}{1-\rho^2}\right)\frac{C_o}{c_1} = 0$$

or

$$L^3 + 3aL + b = 0 \qquad (2.31)$$

where $$a = \frac{\rho^2}{1-\rho^2} \quad \text{and} \quad b = -\frac{2C_o\rho^2}{c_1(1-\rho^2)}$$

Using theory of equations the roots of the cubic equation (2.31) are given by

$$L = p^{1/3}+q^{1/3} \tag{2.32}$$

where $p = \frac{1}{2}\left[-b+ \sqrt{b^2+4a^3}\ \right]$ and $q = \frac{1}{2}\left[-b- \sqrt{b^2+4a^3}\ \right]$

Substituting the value of p and q in terms if ρ, C_o and c_1 in (2.32) on simplification we get the optimum value of L as

$$L^* = \left(\frac{\rho^2}{1-\rho^2}\right)^{1/3}\left[\left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3} + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3}\right] \tag{2.33}$$

Substitution of this value of L in (2.29) gives the optimum value of n as

$$n^* = \frac{C_o}{c_2} - \frac{c_1}{c_2}\left(\frac{\rho^2}{1-\rho^2}\right)^{1/3}\left[\left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3} + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3}\right] \tag{2.34}$$

The above values of n^* and L^* will satisfy the constraint $n-2L\geq 0$ if and only if

$$\{\text{R.H.S. of (2.34)}\} \geq 2\{\text{R.H.S. of (2.33)}\}$$

$$\text{or,} \quad \frac{C_o}{c_2} - \frac{c_1}{c_2}\left(\frac{\rho^2}{1-\rho^2}\right)^{1/3}\left[\left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)^{1/3} + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)^{1/3}\right]$$

$$\geq 2\left(\frac{\rho^2}{1-\rho^2}\right)^{1/3}\left[\left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)^{1/3} + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)\right]$$

$$\text{or,} \quad \frac{C_o}{c_2} \geq \frac{c_1+2c_2}{c_2}\left(\frac{\rho^2}{1-\rho^2}\right)^{1/3}\left[\left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)^{1/3} + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)^{1/3}\right]$$

$$\text{or,} \quad \frac{C_o}{(c_1+2c_2)}\left(\frac{1-\rho^2}{\rho^2}\right)^{1/3} \geq \left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)^{1/3} + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2) + (\rho^2/(1-\rho^2))}\right)^{1/3} \quad (2.35)$$

Raising both sides of inequality (2.35) to the power 3 we get

$$\frac{C_o^3}{(c_1+2c_2)^3}\left(\frac{1-\rho^2}{\rho^2}\right) \geq \frac{2C_o}{c_1} + 3\left(-\frac{\rho^2}{1-\rho^2}\right)^{1/3}\left[\left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3}\right.$$

$$\left. + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3}\right]$$

or,

$$\frac{C_o^3}{(c_1+2c_2)^3}\left(\frac{1-\rho^2}{\rho^2}\right) - \frac{2C_o}{c_1} \geq -3\left(\frac{\rho^2}{1-\rho^2}\right)^{1/3}\left[\left(\frac{C_o}{c_1} + \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3}\right.$$

$$\left. + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3}\right]$$

or, $$\left[\frac{2C_o}{c_1} - \frac{C_o^3}{(c_1+2c_2)^3}\left(\frac{1-\rho^2}{\rho^2}\right)\right]\cdot\frac{1}{3}\left(\frac{1-\rho^2}{\rho^2}\right)^{1/3}$$

$$\leq \left(\frac{C_o}{c_1}+\sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3} + \left(\frac{C_o}{c_1} - \sqrt{(C_o^2/c_1^2)+(\rho^2/(1-\rho^2))}\right)^{1/3} \tag{2.36}$$

Comparing inequalities (2.35) and (2.36), we get

$$\frac{C_o}{(c_1+2c_2)}\left(\frac{1-\rho^2}{\rho^2}\right)^{1/3} \geq \frac{1}{3}\left[\frac{2C_o}{c_1} - \frac{C_o^3}{(c_1+2c_2)^3}\left(\frac{1-\rho^2}{\rho^2}\right)\right]\left(\frac{1-\rho^2}{\rho^2}\right)^{1/3}$$

or,

$$\frac{3C_o}{c_1+2c_2} \geq \frac{2C_o}{c_1} - \frac{C_o^3}{(c_1+2c_2)^3}\left(\frac{1-\rho^2}{\rho^2}\right)$$

which on simplification gives

$$\frac{(c_1+2c_2)^2(4c_2-c_1)}{C_o^2 c_1} \leq \frac{1-\rho^2}{\rho^2} \qquad (2.37)$$

Thus in case the given values of C_o, c_1, c_2 and ρ obey (2.37) the values of L and n given by (2.33) and (2.34) will solve the MPP-II completely.

When (2.37) is not satisfied, that is,

$$\frac{(c_1+2c_2)^2(4c_2-c_1)}{C_o^2 c_1} > \frac{1-\rho^2}{\rho^2} \qquad (2.38)$$

we have to consider the constraint $n-2L \geq 0$ also.

Then the function θ is now defined as:

$$\theta(n,L,u_1,u_2) = -\frac{S_y^2}{n}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] + u_1(C_o - c_1L - c_2n) + u_2(n-2L) \qquad (2.39)$$

The K-T conditions for MPP (2.25) are:

$$\nabla_{(n,L)}\theta = \begin{bmatrix} \frac{S_y^2}{n^2}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] - u_1c_2 + u_2 \\ \frac{2S_y^2\rho^2}{nL^3} - u_1c_1 - 2u_2 \end{bmatrix} \geq \underline{0} \quad (a)$$

$$(n,L)\nabla_{(n,L)}\theta = n\left[\frac{S_y^2}{n^2}\left\{\frac{\rho^2}{L^2} + (1-\rho^2)\right\} - u_1c_2 + u_2\right] + L\left[\frac{2S_y^2\rho^2}{nL^3} - u_1c_1 - 2u_2\right] = 0 \quad \text{(b)} \qquad (2.40)$$

and

$$\begin{pmatrix} n \\ L \end{pmatrix} \geq \underline{0} \quad \text{(c)}$$

$$\nabla_{(u_1,u_2)}\theta = \begin{bmatrix} C_o - c_1L - c_2n \\ n-2L \end{bmatrix} \geq \underline{0} \quad \text{(d)}$$

$$(u_1,u_2)\nabla_{(u_1,u_2)}\theta = u_1(C_o - c_1L - c_2n) + u_2(n-2L) = 0 \quad \text{(e)}$$

and

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \geq 0 \quad \text{(f)}$$

If u_1 and u_2 are $\neq 0$, in order to satisfy (2.40(e)) we must have

$$C_o - c_1L - c_2n = 0$$

and

$$n-2L = 0.$$

solving the above equations for L and n we get

$$L^* = \frac{C_o}{c_1+2c_2} \qquad (2.41)$$

and

$$n^* = \frac{2C_o}{c_1+2c_2} \qquad (2.42)$$

The values of L and n, given in the equation (2.41) and (2.42) respectively will be optimum if and only if u_1 and u_2 are

positive.

Substituting the values of L and n given by (2.41) and (2.42) in (2.40(a)) and taking equality sign we get the following two equations for u_1 and u_2

$$\frac{S_y^2}{\left(\frac{4C_o^2}{(c_1+2c_2)^2}\right)}\left[\frac{\rho^2}{\left(\frac{C_o^2}{(c_1+2c_2)^2}\right)}+(1-\rho^2)\right]-u_1c_2+u_2 = 0$$

or,

$$\frac{S_y^2(c_1+2c_2)^2}{4C_o^2}\left[\frac{\rho^2(c_1+2c_2)^2}{C_o^2}+(1-\rho^2)\right]-u_1c_2+u_2 = 0 \qquad (2.43)$$

and

$$\frac{2S_y^2\rho^2}{\left(\frac{2C_o}{(c_1+2c_2)}\right)\left(\frac{C_o^3}{(c_1+2c_2)^3}\right)} - u_1c_1-2u_2 = 0$$

or
$$\frac{S_y^2\rho^2(c_1+2c_2)^4}{C_o^4} -u_1c_1-2u_2 = 0 \qquad (2.44)$$

Putting $a = \frac{S_y^2(c_1+2c_2)^4}{C_o^4}$ in (2.43) and (2.44), we get

$$\frac{a\rho^2}{4} + \frac{S_y^2\sqrt{a}\ (1-\rho^2)}{4} -u_1c_2+u_2 = 0 \qquad (2.45)$$

and
$$a\rho^2-u_1c_1-2u_2 = 0 \qquad (2.46)$$

Multiplying (2.45) by 2 and adding to (2.46), we get

$$\frac{3a\rho^2}{2} + \frac{S_y\sqrt{a}\ (1-\rho^2)}{2} - 2u_1c_2 - u_1c_1 = 0$$

or,

$$u_1(2c_2+c_1) = \frac{3a\rho^2+S_y\sqrt{a}\ (1-\rho^2)}{2}$$

$$u_1 = \frac{3a\rho^2+S_y\sqrt{a}\ (1-\rho^2)}{2(c_1+2c_2)} > 0 \qquad (2.47)$$

Substitution of the value of u_1 given by (2.47) in (2.46) gives

$$a\rho^2 - c_1\left[\frac{3a\rho^2+S_y\sqrt{a}\ (1-\rho^2)}{2(c_1+2c_2)}\right] - 2u_2 = 0$$

$$\Rightarrow \qquad u_2 = \frac{1}{2}\left[a\rho^2 - \frac{c_1(3a\rho^2+S_y\sqrt{a}\ (1-\rho^2))}{2(c_1+2c_2)}\right] \qquad (2.48)$$

Thus $u_2>0$ if and only if

$$a\rho^2 - \frac{c_1(3a\rho^2+S_y\sqrt{a}\ (1-\rho^2))}{2(c_1+2c_2)} > 0$$

or, $\quad 2a\rho^2(c_1+2c_2) - c_1(3a\rho^2+S_y\sqrt{a}\ (1-\rho^2)) > 0$

or, $\quad 2\rho^2(c_1+2c_2) - 3c_1\rho^2 > \frac{S_y(1-\rho^2)c_1}{\sqrt{a}}$

$$\text{or,} \qquad \rho^2(2c_1+4c_2-3c_1) > \frac{S_y(1-\rho^2)c_1}{\sqrt{a}}$$

$$\text{or,} \qquad \rho^2(4c_2-c_1) > S_y(1-\rho^2)c_1 \cdot \frac{C_o^2}{S_y(c_1+2c_2)^2}$$

$$\text{or} \qquad \frac{(c_1+2c_2)^2(4c_2-c_1)}{C_o^2 c_1} > \frac{1-\rho^2}{\rho^2}$$

which is nothing but condition (2.38).

Thus u_1 and u_2 both are greater than zero under condition (2.38), hence the optimal solution is given by (2.41) and (2.42).

It can be seen that if $\frac{(c_1+2c_2)^2(4c_2-c_1)}{C_o^2 c_1} = \frac{1-\rho^2}{\rho^2}$ the expressions (2.33) and (2.41) give the same value of L^*. similarly expressions (2.34) and (2.42) gives the same value of n^*.

Problem 3 (MPP-III): The MPP-III can also be solved exactly as MPP-II.

Consider MPP-III as:

$$\left.\begin{aligned} &\text{Maximize} && V''(n,L) = -\frac{S_y^2}{n[aL^2+bL+c]} \\ &\text{Subject to} && C_o - c_1 L - c_2 n \geq 0 \\ & && n-2L \geq 0 \\ & && n \geq 0, \quad L \geq 0. \end{aligned}\right\} \qquad (2.49)$$

Ignoring the constraint $n-2L\geq 0$ and using K-T conditions stated in with

$$\theta(n,L,u) = -\frac{S_y^2}{n[aL^2+bL+c]} + u(C_o-c_1L-c_2n)$$

we get the optimum values of L and n as

$$L^* = \frac{(aC_o-bc_1)+\sqrt{(aC_o-bc_1)^2+3ac_1(bC_o-cc_1)}}{3ac_2} \qquad (2.50)$$

and

$$n^* = \frac{C_o}{c_2} - \frac{(aC_o-bc_1)+\sqrt{(aC_o-bc_1)^2+3ac_1(bC_o-cc_1)}}{3ac_2} \qquad (2.51)$$

As before these values of L^* and n^* will satisfy the constraint $n-2L\geq 0$ if and only if

$$[\text{R.H.S. of (2.51)}] \geq 2[\text{R.H.S. of (2.50)}]$$

which on simplification gives

$$\frac{C_o^2(c_1-4c_2)}{c_1+2c_2} \geq \frac{(2bC_oc_2-bC_oc_1-cc_1^2-2cc_1c_2)}{a}, \quad a\neq 0 \qquad (2.52)$$

Thus if (2.52) holds (2.50) and (2.51) will solve MPP-III completely.

In case (2.52) is not satisfied that is if

$$\frac{C_o^2(c_1-4c_2)}{c_1+2c_2} < \frac{(2bC_oc_2-bC_oc_1-cc_1^2-2cc_1c_2)}{a}, \quad a\neq 0 \qquad (2.53)$$

as discussed earlier, both the constraints will become active, that is, we have $u_1\neq 0$, $u_2\neq 0$ and

$$C_o-c_1L-c_2n = 0$$

and

$$n-2L = 0$$

and the optimal values of L and n are as given by (2.41) and (2.42). It can be verified that we can find $u_1,u_2>0$ which along with (2.41), (2.42) and (2.53) satisfy all the K-T conditions for MPP-III.

It can also be seen that when equality holds in (2.53) the expressions (2.41) and (2.50) give the same value of L^* and expressions (2.42) and (2.51) give the same value of n^*.

CHAPTER-III

DETERMINATION OF THE OPTIMUM STRATA BOUNDARIES

3.1 INTRODUCTION

When a single characteristic is under study the best criteria for stratification is the frequency distribution of the characteristic itself. The next best is the frequency distribution of some other variable (known as stratification variable) which is highly correlated with the study variable. Dalenius (1957) worked out the optimum stratum boundaries under proportional and Neyman allocations when the number of strata L is known and the study variable itself is used as stratification variable. Some earlier studies are due to Dalenius (1950), Dalenius, T and Gurney (1951), Mahalanobis (1952), Hansen Hurwitz and Madow (1953), Aoyama (1954). Later works are due to Dalenius and Hodges (1959), Durbin (1959), Sethi (1963), Murth (1967) and several others. Hess, Sethi and Balakrishnan (1966) made a comparison of various approximate methods of stratification.

Most of these authors obtained calculus equations for the strata boundaries which are ill adapted to practical computations. They obtained only approximate solutions under certain assumptions.

Unnithan (1978) suggested an iterative procedure for obtaining the optimum strata boundaries by minimizing the variance under Neyman allocation using Modified Newton Method. He showed that the iterative procedure of Dalenius and Hodges (1959) is slow in attaining even a local minimum. The procedure may oscillate and does nct suggest any stopping rule. Both the methods requires initial approximate solution using cumulative $\sqrt{f}$ rule. Also there is no guarantee that the procedure provide a global minimum in the absence of a suitable approximate initial solution and the variance function has more than one local minima.

In this chapter the problem of determining the optimum stratum boundaries, when the number of strata L is known and Neyman allocation is used, is formulated as a mathematical programming problem (MPP). A solution procedure is developed using dynamic programming technique which provides the global minimum of the objective function in a finite number of steps. This chapter is based on my joint research paper "Optimum stratification: A Mathematical Programming Approach", accepted for presentation in the VI Islamic Society of Statistical Sciences Conference (ISOSSC) to be held in Dhaka during December 12-15, 1998.

3.2 FORMULATION OF THE PROBLEM

Let f(x) denote the frequency function of the continuous study variable x, $x_o \leq x \leq x_L$ where x_o and x_L are known real numbers and $x_o < x_L$.

The problem of constructing L strata between x_o and x_L can be considered as the problem of determining the L-1 stratification points $x_1, x_2, \ldots, x_{L-1}$ such that the sampling variance of the stratified sample mean $\bar{x}_{st}$ is minimum. Where $\bar{x}_{st}$ is the usual estimator of the overall population mean $\bar{X}$.

To build a mathematical model for the above problem it is necessary to express the variance of $\bar{x}_{st}$ as a function of the stratification points x_h; h=1,2,...,L-1.

Ignoring the finite population correction (f.p.c.) the variance of $\bar{x}_{st}$ under Neyman allocation is given by

$$V(\bar{x}_{st}) = \frac{\left(\sum_{h=1}^{L} W_h \sigma_h \right)^2}{n} ,$$

where W_h and σ_h^2 are the stratum weight and stratum variance for the h^{th} stratum; h=1,2,...,L respectively and n is the known fixed total sample size. In order to minimize $V(\bar{x}_{st})$ it is sufficient to minimize $\sum_{h=1}^{L} W_h \sigma_h$ only, because n is a known constant.

As the study variable x is assumed to be continuous we have

$$W_h = \int_{x_{h-1}}^{x_h} f(x)\,dx \qquad (3.1)$$

and $$\sigma_h^2 = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} x^2 f(x)\,dx - \mu_h^2. \qquad (3.2)$$

Where

$$\mu_h = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} x f(x)\,dx \qquad (3.3)$$

is the stratum mean of the h^{th} stratum; h=1,2,...,L.

Using (3.1), (3.2) and (3.3) we can express $W_h\sigma_h$ as a function of x_h and x_{h-1} only.

Let $$f_h(x_h, x_{h-1}) = W_h\sigma_h \qquad (3.4)$$

The problem of determining the Optimum Strata Boundaries (OSB) may be stated as:

"Find $x_1, x_2, \ldots, x_{L-1}$ which minimize $\sum_{h=1}^{L} f_h(x_h, x_{h-1})$, subject to the constraint $x_o \le x_1 \le x_2 \le \ldots \le x_{L-1} \le x_L$.

The problem of determining the OSB could further be simplified by defining

$$y_h = x_h - x_{h-1};\ h=1,2,\ldots,L \qquad (3.5)$$

where $y_h \ge 0$ denotes the width of the h^{th} stratum.

From (3.5)
$$\sum_{h=1}^{L} y_h = \sum_{h=1}^{L} (x_h - x_{h-1}) = d \qquad (3.6)$$

where $d=x_L-x_o$.

The above problem can thus be stated as the following Mathematical Programming Problem (MPP)

$$\left.\begin{aligned} &\text{Minimize} && \sum_{h=1}^{L} f_h(x_h, x_{h-1}) \\ &\text{subject to} && \sum_{h=1}^{L} y_h = d \\ &\text{and} && y_h \geq 0;\ h=1,2,\ldots,L. \end{aligned}\right\} \qquad (3.7)$$

The k^{th} stratification point x_k; $k=1,2,\ldots,L-1$ can be expressed as a function of $y_1, y_2, \ldots, y_k$ as

$$x_k = x_o + y_1 + y_2 + \ldots + y_k$$

$$x_k - x_o = \sum_{h=1}^{k} y_h = d_k$$

where d_k is the total width available for division into k strata.

Since x_o is known, the first term $f_1(y_1, x_o) = f_1(x_o+y_1, x_o)$ in the objective function of the MPP (3.7) is a function of y_1 alone. Once y_1 is known the next stratification point $x_1=x_o+y_1$ will be known and the second term in the objective function $f_2(y_2, x_1)$ will

become a function of y_2 alone. Due to this special feature of the objective function and the separable nature of the constraint function, dynamic programming technique may be used to solve the MPP (3.7).

Writing the objective function as a function of y_h alone we get the MPP (3.7) as:

$$\left.\begin{aligned} &\text{Minimize} && \sum_{h=1}^{L} f_h(y_h) \\ &\text{Subject to} && \sum_{h=1}^{L} y_h = d, \\ &\text{and} && y_h \geq 0;\ h=1,2,\ldots,L. \end{aligned}\right\} \qquad (3.8)$$

3.3 THE SOLUTION USING DYNAMIC PROGRAMMING TECHNIQUE

The problem (3.8) is a multistage decision problem in which the objective function and the constraints are separable functions of y_h.

A function of n variables $f(x_1,x_2,\ldots,x_n)$ is called separable if it can expressed as

$$f(x_1,x_2,\ldots,x_n) = f_1(x_1)+f_2(x_2)+\ldots+f_n(x_n) = \sum_{j=1}^{n} f_j(x_j).$$

Apart from the separability, the MPP (3.8) also possesses other features stated in section 1.9 of Chapter-I, allowing us the use of dynamic programming technique.

Consider the following subproblem of (3.8) for k(<L) strata.

$$\left.\begin{aligned} &\text{Minimize} && \sum_{h=1}^{k} f_h(y_h) \\ &\text{Subject to} && \sum_{h=1}^{k} y_h = d_k, \\ &\text{and} && y_h \geq 0;\ h=1,2,\ldots,k \end{aligned}\right\} \qquad (3.9)$$

where as defined earlier $d_k < d$ is the total width available for division into k strata.

Note that $d_k = d$ for k=L

Also

$$\begin{aligned} d_k &= y_1+y_2+\ldots+y_k \\ d_{k-1} &= y_1+y_2+\ldots+y_{k-1} = d_k-y_k \\ d_{k-2} &= y_1+y_2+\ldots+y_{k-2} = d_{k-1}-y_{k-1} \\ &\vdots \qquad\qquad\qquad\quad \vdots \\ d_2 &= y_1+y_2 \qquad\qquad = d_3-y_3 \\ d_1 &= y_1 \qquad\qquad\quad = d_2-y_2 \end{aligned}$$

Let $f(k,d_k)$ denotes the minimum value of the objective function of (3.9), that is,

$$f(k,d_k) = \min\left[\sum_{h=1}^{k} f_h(y_h) \;\middle|\; \sum_{h=1}^{k} y_h=d_k \text{ and } y_h\geq 0;\ h=1,2,\ldots,k\right].$$

With the above definition of $f(k,d_k)$ the problem (3.8) is equivalent to finding $f(L,d)$ recursively by finding $f(k,d_k)$ for $k=1,2,\ldots,L$ and $0\leq d_k\leq d$.

We can write

$$f(k,d_k) = \min\left[f_k(y_k)+\sum_{h=1}^{k-1} f_h(y_h) \;\middle|\; \sum_{h=1}^{k-1} y_h=d_k-y_k \text{ and } y_h\geq 0;\ h=1,2,\ldots,k\right]$$

For a fixed value of y_k; $0\leq y_k\leq d_k$

$$f(k,d_k) = f_k(y_k)+\min\left[\sum_{h=1}^{k-1} f_h(y_h) \;\middle|\; \sum_{h=1}^{k-1} y_h=d_k-y_k \text{ and } y_h\geq 0;\ h=1,2,\ldots,k-1\right]$$

Using the Bellman's principle of optimality stated in section 1.9 of Chapter-I we get the recurrence relation of the Dynamic Programming as

$$f(k,d_k) = \min_{0\leq y_k\leq d_k}\left[f_h(y_h)+f(k-1,d_k-y_k)\right], \text{ for } k\geq 2 \qquad (3.10)$$

For the first stage, that is for $k=1$

$$f(1,d_1) = f_1(d_1) \Rightarrow y_1^* = d_1 \qquad (3.11)$$

where y_1^* is the optimum width of the first stratum.

The relations (3.11) and (3.10) are solved recursively for each $k=1,2,\ldots,L$ and $0 \le d_k \le d$ and $f(L,d)$ is obtained. From $f(L,d)$ the optimum width of L^{th} stratum, y_L^*, is obtained; from $f(L-1,d-y_L^*)$ the optimum width of $(L-1)^{th}$ stratum, y_{L-1}^*, is obtained and so on until y_1^* is obtained.

3.4 NUMERICAL ILLUSTRATIONS

Example 1: Let x follows Uniform Distribution within the interval [a,b].

Then $$f(x) = \frac{1}{b-a} \; ; \; a \le x \le b$$

$$= 0 \quad ; \text{ otherwise.}$$

Using (3.1), the stratum weight W_h for the h^{th} stratum is given as

$$W_h = \int_{x_h-1}^{y_h+x_{h-1}} f(x)\,dx = \int_{x_{h-1}}^{y_h+x_{h-1}} \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}\left[x\right]_{x_{h-1}}^{y_h+x_{h-1}}$$

$$= \frac{y_h+x_{h-1}-x_{h-1}}{b-a}$$

or $$W_h = \frac{y_h}{b-a} \qquad (3.12)$$

Using (3.3) the stratum mean μ_h for the h^{th} stratum is given as

$$\mu_h = \frac{1}{W_h}\int x f(x)\,dx$$

$$= \frac{1}{W_h}\int \frac{x}{b-a}\,dx$$

$$= \frac{1}{W_h(b-a)}\left[\frac{x^2}{2}\right]_{x_{h-1}}^{y_h+x_{h-1}}$$

$$= \frac{(y_h+x_{h-1})^2 - x_{h-1}^2}{2(b-a)W_h}$$

$$= \frac{(y_h+x_{h-1}+x_{h-1})(y_h+x_{h-1}-x_{h-1})}{2(b-a)W_h}$$

$$= \frac{(y_h+2x_{h-1})y_h}{2(b-a)} \times \frac{(b-a)}{y_h} \qquad (\text{By } (3.10))$$

or $$\mu_h = \frac{y_h+2x_{h-1}}{2}\,. \qquad (3.13)$$

Using (3.2), (3.12) and (3.13) the stratum variance of the h^{th} stratum is given as

$$\sigma_h^2 = \frac{1}{W_h}\int x^2 f(x)\,dx - \mu_h^2 = \frac{1}{W_h(b-a)}\left[\frac{x^3}{3}\right]_{x_{h-1}}^{y_h+x_{h-1}} - \mu_h^2$$

$$= \frac{(y_h+x_{h-1})^3-x_{h-1}^3}{W_h\,3(b-a)} - \mu_h^2$$

$$= \frac{(y_h+x_{h-1}-x_{h-1})\left[(y_h+x_{h-1})^2+(y_h+x_{h-1})x_{h-1}+x_{h-1}^2\right]}{W_h\,3(b-a)} - \mu_h^2$$

$$= \frac{y_h(y_h^2+2y_hx_{h-1}+x_{h-1}^2+y_hx_{h-1}+x_{h-1}^2+x_{h-1}^2}{3(b-a)} \times \frac{b-a}{y_h} - \frac{(y_h+2x_{h-1})^2}{4}$$

$$= \frac{y_h^2+3y_hx_{h-1}+3x_{h-1}^2}{3} - \frac{y_h^2+4y_hx_{h-1}+4x_{h-1}^2}{4}$$

$$= \frac{4y_h^2+12y_hx_{h-1}+12x_{h-1}^2-3y_h^2-12y_hx_{h-1}-12x_{h-1}^2}{12}$$

or
$$\sigma_h^2 = \frac{y_h^2}{12} \qquad (3.14)$$

Using (3.4) and the values of W_h and σ_h^2 given by (3.12) and (3.14) respectively the objective function of the MPP (3.8) may be expressed as a function of y_h alone as:

$$\sum_{h=1}^{L} f(x_h, x_{h-1}) = \sum_{h=1}^{L} \left(\frac{y_h}{b-a}\right)\sqrt{\frac{y_h^2}{12}}$$

$$= \sum_{h=1}^{L} \frac{y_h^2}{2\sqrt{3}\,(b-a)}$$

$$= \sum_{h=1}^{L} f(y_h).$$

Thus MPP (3.8) takes the form:

$$\left.\begin{aligned} &\text{Minimize} && \sum_{h=1}^{L} \frac{y_h^2}{2\sqrt{3}\,d} \\ &\text{Subject to} && \sum_{h=1}^{L} y_h = d, \\ &\text{and} && y_h \geq 0;\ h=1,2,\ldots,L \end{aligned}\right\} \qquad (3.15)$$

where d=b-a.

To illustrate the computational procedure let [a,b]=[1,2] and L=6. This gives MPP (3.15) as:

$$\left.\begin{aligned} &\text{Minimize} && \sum_{h=1}^{6} \frac{y_h^2}{2\sqrt{3}} \\ &\text{Subject to} && \sum_{h=1}^{6} y_h = 1, \\ &\text{and} && y_h \geq 0;\ h=1,2,\ldots,6. \end{aligned}\right\} \qquad (3.16)$$

Using the recurrence relations (3.10) and (3.11) for the MPP (3.16) we get: For the first stage (k=1)

$$f(1,d_1) = \frac{d_1^2}{2\sqrt{3}} \quad \text{at } y_1^* = d_1 \tag{3.17}$$

For the second stage (k=2)

$$f(2,d_2) = \min_{0 \le y_2 \le d_2} \left[\frac{y_2^2}{2\sqrt{3}} + f(1,d_2-y_2) \right]$$

$$= \min_{0 \le y_2 \le d_2} \left[\frac{y_2^2}{2\sqrt{3}} + \frac{(d_2-y_2)^2}{2\sqrt{3}} \right] \tag{3.18}$$

Differentiating the quantity inside [ ] on the right hand side of (3.18) with respect to y_2 and equating to zero we get

$$\frac{1}{2\sqrt{3}} \left[4y_2 + 2d_2 \right] = 0 \quad \text{or } y_2 = \frac{d_2}{2}$$

which implies that

$$f(2,d_2) = \frac{d_2^2}{4\sqrt{3}} \quad \text{at} \quad y_2^* = \frac{d_2}{2} \; . \tag{3.19}$$

For the third stage (k=3)

$$f(3,d_3) = \min_{0 \le y_3 \le d_3} \left[\frac{y_3^2}{2\sqrt{3}} + f(2,d_3-y_3) \right]$$

$$= \min_{0 \le y_3 \le d_3} \left[\frac{y_3^2}{2\sqrt{3}} + \frac{(d_3 - y_3)^2}{4\sqrt{3}} \right] \qquad (3.20)$$

As before, using differential calculus for minimization of the quantity inside [ ] in (3.20), we get

$$f(3,d_3) = \frac{d_3^2}{6\sqrt{3}} \quad \text{at} \quad y_3^* = \frac{d_3}{3} \qquad (3.21)$$

Similarly for the fourth and fifth stages we get

$$f(4,d_4) = \frac{d_4^2}{8\sqrt{3}} \quad \text{at} \quad y_4^* = \frac{d_4}{4} \qquad (3.22)$$

$$f(5,d_5) = \frac{d_5^2}{10\sqrt{3}} \quad \text{at} \quad y_5^* = \frac{d_5}{5} \qquad (3.23)$$

For the final stage (k=6)

$$f(6,d_6) = \min_{0 \le y_6 \le d_6} \left[\frac{y_6^2}{2\sqrt{3}} + \frac{(d_6 - y_6)^2}{10\sqrt{3}} \right]$$

or $$f(6,1) = \min_{0 \le y_6 \le 1} \left[\frac{y_6^2}{2\sqrt{3}} + \frac{(1-y_6)^2}{10\sqrt{3}} \right],$$

because we have only six strata, d_6=b-a=2-1=1.

Now

$$f(6,1) = \frac{1}{12\sqrt{3}} = 0.048112522$$

at $y_6^* = \frac{1}{6} = 0.166667$ (3.24)

From (3.24), $d_5 = d_6 - y_6^* = 1 - 0.166667 = 0.833333$

Substituting this value of d_5 in (3.23) we get

$$y_5^* = \frac{0.833333}{5} = 0.166666$$

Proceeding in the same manner, we get

$$y_4^* = 0.166667,$$

$$y_3^* = 0.166666,$$

$$y_2^* = 0.166667$$

and $$y_1^* = 0.166667.$$

The optimum strata boundaries are then obtained as

$$x_1^* = x_o + y_1^* = 1 + 0.166667 = 1.166667$$

$$x_2^* = x_1^* + y_2^* = 1.166667 + 0.166667 = 1.333334$$

$$x_3^* = x_2^* + y_3^* = 1.333334 + 0.166666 = 1.500000$$

$$x_4^* = x_3^* + y_4^* = 1.500000 + 0.166667 = 1.666667$$

$$x_5^* = x_4^* + y_5^* = 1.666667 + 0.166666 = 1.833333$$

The optimum value of the objective function is:

$$\sum_{h=1}^{6} f_h(y_h) = f(6,1) = 0.048112522.$$

Example 2: Let x follows the Right Triangular distribution in the interval [a,b]. Then

$$f(x) = \frac{2(b-x)}{(b-a)^2} \quad ; \quad a \le x \le b$$

$$= 0 \quad ; \text{ otherwise.}$$

By (3.1) we have

$$W_h = \int_{x_h - 1}^{y_h + x_{h-1}} \frac{2}{(b-a)^2} (b-x)\,dx = \frac{2}{(b-a)^2} \left[bx - \frac{x^2}{2} \right]_{x_{h-1}}^{y_h + x_{h-1}}$$

$$= \frac{1}{(b-a)^2} \left[2bx - x^2 \right]_{x_{h-1}}^{y_h + x_{h-1}}$$

$$= \frac{2by_h + 2bx_{h-1} - y_h^2 - x_{h-1}^2 - 2y_h x_{h-1} - 2bx_{h-1} + x_{h-1}^2}{(b-a)^2}$$

$$= \frac{y_h (2b - 2x_{h-1} - y_h)}{(b-a)^2}$$

$$= \frac{y_h[2(b-x_{h-1})-y_h]}{(b-a)^2}$$

$$\left.\begin{aligned} \text{or} \quad W_h &= \frac{y_h(2a_h-y_h)}{(b-a)^2}, \\ \text{where} \quad a_h &= b-x_{h-1}. \end{aligned}\right\} \qquad (3.25)$$

Again (3.3) gives

$$\mu_h = \frac{1}{W_h}\int_{x_h-1}^{y_h+x_{h-1}} xf(x)\,dx$$

$$= \frac{2}{W_h(b-a)^2}\int_{x_h-1}^{y_h+x_{h-1}} x(b-x)\,dx$$

$$= \frac{2}{W_h(b-a)^2}\left[b\frac{x^2}{2}-\frac{x^3}{3}\right]_{x_{h-1}}^{y_h+x_{h-1}}$$

$$= \frac{2}{6W_h(b-a)^2}\left[3bx^2-2x^3\right]_{x_{h-1}}^{y_h+x_{h-1}}$$

$$= \frac{2}{6(b-a)^2}\cdot\frac{(b-a)^2}{y_h(2a_h-y_h)}\left[3b(y_h+x_{h-1})^2-2(y_h+x_{h-1})^3-3bx_{h-1}^2+2x_{h-1}^3\right]$$

$$= \frac{1}{3y_h(2a_h-y_h)}\left[3b(y_h+2x_{h-1})y_h-2\{y_h((y_h+x_{h-1})^2+(y_h+x_{h-1})x_{h-1}+x_{h-1}^2)\}\right]$$

$$= \frac{3b(y_h+2x_{h-1}) - 2(y_h^2+2x_{h-1}y_h+x_{h-1}^2+y_h x_{h-1}+x_{h-1}^2+x_{h-1}^2)}{3(2a_h-y_h)}$$

or
$$\mu_h = \frac{3b(y_h+2x_{h-1}) - 2(y_h^2+3x_{h-1}y_h+3x_{h-1}^2)}{3(2a_h-y_h)} \qquad (3.26)$$

Using (3.2) we get

$$\sigma_h^2 = \frac{1}{W_h}\int_{x_h-1}^{y_h+x_{h-1}} x^2 f(x)\,dx - \mu_h^2 = \frac{2}{W_h(b-a)^2}\int_{x_h-1}^{y_h+x_{h-1}} x^2(b-x)\,dx - \mu_h^2$$

$$= \frac{2}{(b-a)^2}\cdot\frac{(b-a)^2}{y_h(2a_h-y_h)}\left[\frac{bx^3}{3} - \frac{x^4}{4}\right]_{x_{h-1}}^{y_h+x_{h-1}} - \mu_h^2$$

$$= \frac{2}{12y_h(2a_h-y_h)}\left[4bx^3-3x^4\right]_{x_{h-1}}^{y_h+x_{h-1}} - \mu_h^2$$

$$= \frac{4b(y_h+x_{h-1})^3 - 3(y_h+x_{h-1})^4 - 4bx_{h-1}^3 + 3x_{h-1}^2}{6y_h(2a_h-y_h)} - \mu_h^2$$

Substituting the value of μ_h from (3.26) in the above expression for σ_h^2 and simplifying we get

$$\sigma_h^2 = \frac{y_h^2(y_h^2-6a_h y_h+6a_h^2)}{18(2a_h-y_h)^2}, \qquad (3.27)$$

where
$$a_h = b-x_{h-1}; \quad h=1,2,\ldots,L. \qquad (3.28)$$

Using (3.4) and the values of W_h and σ_h^2 given by (3.25) and (3.27) respectively, the objective function of the MPP (3.8) may be expressed as a function of y_h alone as:

$$\sum_{h=1}^{L} f(x_h, x_{h-1}) = \sum_{h=1}^{L} \frac{y_h(2a_h - y_h)}{(b-a)^2} \sqrt{\frac{y_h^2(y_h^2 - 6a_h y_h + 6a_h^2)}{18(2a_h - y_h)^2}}$$

$$= \sum_{h=1}^{L} \left(\frac{y_h^2}{3\sqrt{2}\,(b-a)^2}\right) \sqrt{y_h^2 - 6a_h y_h + 6a_h^2}$$

$$= \sum_{h=1}^{L} f(y_h)$$

because a_h, by definition (3.28), is a function of x_{h-1} and $x_{h-1} = -x_{h-2} + y_{h-1}$ is known before the value of y_h is determined.

Thus the MPP (3.8) for example-2 takes the form:

$$\left.\begin{aligned} &\text{Minimize} \quad \sum_{h=1}^{L} \frac{y_h^2 \sqrt{y_h^2 - 6a_h y_h + 6a_h^2}}{3\sqrt{2}\,(b-a)^2} \\ &\text{subject to} \quad \sum_{h=1}^{L} y_h = d, \\ &\text{and} \quad y_h \geq 0;\ h = 1, 2, \ldots, L \end{aligned}\right\} \qquad (3.29)$$

where d=b-a.

To illustrate the computational procedure let [a,b]=[0,1] and L=6. This gives MPP (3.29) as:

$$\left.\begin{aligned} &\text{Minimize} \quad \sum_{h=1}^{6} \frac{y_h^2\sqrt{y_h^2-6a_hy_h+6a_h^2}}{3\sqrt{2}} \\ &\text{subject to} \quad \sum_{h=1}^{6} y_h = 1, \\ &\text{and} \quad y_h \geq 0;\ h=1,2,\ldots,L. \end{aligned}\right\} \qquad (3.30)$$

where $a_h=1-x_{h-1}$; $h=1,2,\ldots,L$.

Using the recurrence relations (3.10) and (3.11) for the MPP (3.30) we get:

For the first stage (k=1)

$$f(1,d_1) = \frac{d_1^2\sqrt{d_1^2-6d_1+6}}{3\sqrt{2}} \text{ at } y_1^*=d_1 \qquad (3.31)$$

and for k^{th} stage, $k\geq 2$,

$$f(k,d_k) = \min_{0\leq y_k\leq d_k}\left[\frac{y_k^2\sqrt{y_k^2-6a_ky_k+6a_k^2}}{3\sqrt{2}} + f(k-1,d_k-y_k)\right] \qquad (3.32)$$

where $a_k=1-x_{k-1}=1-(x_o+y_1+y_2+\ldots+y_{k-1})=1-(y_1+y_2+\ldots+y_{k-1})=1-d_{k-1}$

$\Rightarrow a_k = 1-(d_k - y_k) = 1 - d_k + y_k$

substituting this value of a_k in (3.32) and executing the computer program, given in the following section, developed for the solution described in section 3.3, the optimum strata widths are obtained as

$$y_1^* = 0.1130,$$

$$y_2^* = 0.1205,$$

$$y_3^* = 0.1305,$$

$$y_4^* = 0.1460,$$

$$y_5^* = 0.1735,$$

$$y_6^* = 0.3165.$$

The optimum value of the objective function is

$$\sum_{h=1}^{6} f_h(y_h) = f(6,1) = 0.0420983170.$$

And the optimum strata boundaries (OSB) are:

$$x_1^* = x_o + y_1^* = 0+0.1130=0.1130$$

$$x_2^* = x_1^* + y_2^* = 0.1130+0.1205=0.2335$$

$$x_3^* = x_2^* + y_3^* = 0.2335+0.1305=0.3640$$

3.5 THE COMPUTER PROGRAMMING

```
/*                          E A KHAN
        To get higher accuracy change the value of gridpt and interadpt.
Accuracy of y is upto the level as the number of digit in interadpt*
gridpt i.e. if gridpt=1000 and interadpt=1000 then accuracy is upto 6
decimal places .If U put gridpt more than 1000 then change values
[1005] in line 13 & 14 . Do not enter stage value more than 9 .If you
want to calculate for higher stage then change the values  [10] in line
no.-14 to higher values */

#include<stdio.h>
#include<math.h>
 int i,j,k,m,m1,m2,p,limit=1,gridpt=1000,interadpt=1000,sootgrid,gridchk,stage
float d[1005],griddif,dif,tmpdif,dd;
double f[10][1005],y[10][1005],n1,x2,yy,fx,fx1,fx2,finaly[10],finald[10];

FILE *fp,*fp1,*fp2;
main()
 {
   /*      fp2=fopen("mrest.res","w");
         fp1=fopen("mamk1.res","w");          */
fp=fopen("mamk.res","a");
  griddif=limit*1.0/gridpt;
  dif=griddif*1.0/interadpt;
  n1=3.0*1.41421356?;
  m=gridpt*interadpt;

  printf(" enter the stage value (1 to 9 only):");
  scanf("%d",&stage);
  if(stage>9)
   { printf(" stage can not be more than 9 ");exit(0);}
  fprintf(fp," \n NUMBER OF STAGE= %d \n",stage);

  for(j=0;j<gridpt;j++)
   d[j]=j*griddif;

  d[gridpt]=limit;

   i=1;
    for(j=0;j<=gridpt;j++)
     fun1();

     for(i=2;i<stage;i++)
     {
       for(j=0;i<=gridpt;j++)
temp();
      }
       i=stage; j=gridpt;
       temp();

/* to get details calculation U can remove /* in line 19,20,51 & 60 */
/*
 for(j=0;j<=gridpt;j++)
         {
   fprintf(fp1,"\n j=%d d=%6.4f ",j,d[j]);
              for(i=1;i<stage;i++)
      fprintf(fp1,"f=%12.9f y=%12.5f ",f[i][j],y[i][j]);


      fprintf(fp1,"\n\n Final Stage:\n  %d   d=1    f=%12.9f
      y=%12.5f ",gridpt,f[stage][gridpt],y[stage][gridpt]);
          */
```

```
/*    Backward calculation to get the final results    */
    i=stage;
    j=gridpt;
    finaly[i]=y[i][j];
    finald[i]=d[j];
      while(i>2)
        {
   i--;
   finald[i]=finald[i+1]-finaly[i+1];
   d[j]=finald[i];
    gridchk=m*finald[i];
    gotgrid=0;
     for(k=0;k<=gridpt;k++)
      {
       m1=m*d[k];
       m2=m*d[k+1];
       if(gridchk==m1)
        { p=k;gotgrid=1;break;}
        else if(gridchk>m1 && gridchk<m2)
         {p=k;break;}
      }
 if(gotgrid==1)
          finaly[i]=y[i][p];
         else
          {
  j=p+1;
  d[j]=finald[i];
  temp();
   finaly[i]=y[i][j];
          }


} /* end of while loop  */
finald[1]=finald[2]-finaly[2];
 finaly[1]=finald[1];
         for(i=1;i<=stage;i++)
   fprintf(fp,"\ny%d = %12.6f  d%d= %6.6f\n",i,finaly[i],i,finald[i]);
  fprintf(fp,"\nf%d=%12.10f\n",i-1,f[stage][gridpt]);

}  /* end of main */


    fun1()
     {
dd=d[j];
yy=dd;
   fx=yy*yy+2.0*(1-dd+yy)*(1.0-dd);
   if(fx<0.0)
    { printf("SQRT OF THE -VE QUANTITY in fun1");
printf("i=%d  j=%d    fx= %f\n",i,j,fx);
     exit(0);}
  else
  fx2=sqrt(fx);
  fx=yy*yy*fx2/n1;
     f[i][j]=fx;
     y[i][j]=yy;

/*   fprintf(fp2,"yy= %f    fx=%f     f[i][j]= %f\n",yy,fx,f[i][j]);
 */

}


      temp()
      {
 dd=d[j];
        f[i][j]=9999.0;
        y[i][j]=9999.0;
        yy=0.0;
        tmpdif=dif;
if(gridpt>100)
 {
   dif=tmpdif;
   if(j>2) dif=griddif;
   if(j>20) dif=10.0*griddif;
   if(j>200) dif=100.0*griddif;
```

```
  }
 else if(j>3)
  dif=griddif;
 else
  dif=tmpdif;
 while(dif>tmpdif)
   {
     fun2();

   yy=y[i][j]-dif;
     if(yy<0.0)
       yy=0.0;
     dd=y[i][j]+dif;
     if(dd>d[j])
       dd=d[j];
     dif=dif/10.0;
   }
   dif=tmpdif;
   fun2();
 /*
fprintf(fp2,"MINIMUM VALUE j=%d d=%f f=%f y=%f \n",j,d[j],f[i][j],y[i][j]);
 */
 }


     fun2()
 {
 while(yy<=dd)
  {
 x2=d[j]-yy;
   gridchk=m*x2;
   gotgrid=0;
    for(k=0;k<=gridpt;k++)
     {
       m1=m*d[k];
       m2=m*d[k+1];
      if(gridchk==m1)
       { p=k;gotgrid=1;break;}
       else if(gridchk>m1 && gridchk<m2)
        {p=k;break;}
     }
 if(gotgrid==1)
 fx1=f[i-1][p];
  else if(i==2)
     fx1=x2*x2*sqrt(x2*x2-6.0*x2+6.0)/n1;

 else
  fx1=f[i-1][p]+(x2-d[p])*(f[i-1][p+1]-f[i-1][p])/griddif;


   fx=yy*yy+6.0*(1-d[j]+yy)*(1-d[j]);
   if(fx<0.0)
    { printf("SQRT OF THE -VE QUANTITY IN FUN2 ");
printf("i=%d j=%d  fx= %f\n",i,j,fx);
     exit(0);}
   else
   x2=sqrt(fx);
   fx=yy*yy*fx2/n1+fx1;
    if(f[i][j]>fx)
     {
      f[i][j]=fx;
      y[i][j]=yy;
            }
yy=yy+dif;
}

   }
```

COMPUTER CENTRE ALIGARH MUSLIM UNIVERSITY

```
RESULTS:-
 NUMBER OF STAGE= 2
y1 =      0.354249   d1= 0.354249
y2 =      0.645751   d2= 1.000000
f2=0.1226262641
 NUMBER OF STAGE= 3
y1 =      0.229791   d1= 0.229791
y2 =      0.272862   d2= 0.502653
y3 =      0.497347   d3= 1.000000
f3=0.0829362599
 NUMBER OF STAGE= 4
y1 =      0.170471   d1= 0.170471
y2 =      0.190530   d2= 0.361001
y3 =      0.226403   d3= 0.587409
y4 =      0.412591   d4= 1.000000
f4=0.0626669993
 NUMBER OF STAGE= 5
y1 =      0.135641   d1= 0.135641
y2 =      0.147360   d2= 0.283001
y3 =      0.164895   d3= 0.447896
y4 =      0.195550   d4= 0.643446
y5 =      0.356554   d5= 1.000000
f5=0.0503620665
 NUMBER OF STAGE= 6
y1 =      0.112647   d1= 0.112647
y2 =      0.120353   d2= 0.233000
y3 =      0.130930   d3= 0.363930
y4 =      0.146071   d4= 0.510001
y5 =      0.173603   d5= 0.683604
y6 =      0.316396   d6= 1.000000
f6=0.0420973209
```

CHAPTER-IV

DETERMINING THE OPTIMUM ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING

4.1 INTRODUCTION

In stratified sample surveys the sample sizes from different strata (called allocations) must be known before drawing the sample. They may be chosen to minimize the sampling variance of the estimator for a prefixed cost of the survey or to minimize the cost for a specified precision of the estimator. When the population mean $\bar{Y}$ of a characteristic y is of interest, it is well known that the sample allocations n_h^* that minimizes the variance

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \qquad (4.1)$$

of the stratified sample mean

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \qquad (4.2)$$

for the prefixed cost

$$C = c_o + \sum_{h=1}^{L} c_h n_h \qquad (4.3)$$

of the survey are given by

$$n_h^* = \frac{(C-c_o)W_h S_h/\sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h/\sqrt{c_h}} \; ; \; h=1,2,\ldots,L \qquad (4.4)$$

Where the population of size N is divided into L non-overlapping and exhaustive strata and for the h^{th} stratum (h=1,2,...,L)

N_h = Stratum size

n_h = Sample size

y_{hi} = Value of the characteristic y for the i^{th} unit

$W_h = \frac{N_h}{N}$ = Stratum weight

$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ = Sample mean

$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$ = Stratum mean

$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y})^2$ = Stratum variance

c_h = Per unit cost of measurement.

Also $$\bar{Y} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} Y_{hi}$$

$$= \frac{1}{N} \sum_{h=1}^{L} N_h \bar{Y}_h$$

$$= \sum_{h=1}^{L} W_h \bar{Y}_h = \text{the overall population mean.}$$

C = the total amount available for the survey

and c_o = overhead cost

In multivariate stratified sampling where more than one characteristics are to be estimated, an allocation which is optimum for one characteristic may not be optimum for other characteristics also. In such situations a compromise criterion is needed to work out a usable allocation which is optimum for all characteristics in some sense. Such an allocation may be called a "Compromise Allocation" because it is based on some compromise criterion.

In surveys where several characteristics defined on the population units are highly correlated, the individual optimum allocations for different characteristics may differ relatively little. For such situations Cochran [1977] suggested the use of the characterwise average of the individual optimum allocations as a usable compromise allocation. He assumed all the characteristics

equally important.

Several others have studied various criteria for obtaining a usable compromise allocation. Among them are Neyman (1934), Peter and Bucher (undated), Geary (1949), Dalenius [1957], Gosh [1958], Yates [1960], Aoyama [1963], Folks and Antle [1965], Kokan and Khan [1967], Chatterji [1967] and [1968], Ahsan and Khan [1977] and [1982], Jahan et al [1994], Khan et al [1997] and many others.

Chaddha et al [1971] used dynamic programming technique to find the optimum allocation in univariate stratified sampling. Omule [1958] used the same technique for the multivariate case. He minimized the total cost of the survey when the tolerance levels for the precisions of the estimates of various characteristics are prefixed.

In this chapter the problem of obtaining a compromise allocation in multivariate stratified random sampling is formulated as a nonlinear mathematical programming problem (NLMPP). This NLMPP is treated as a multistage decision problem and a solution procedure is developed using the dynamic programming technique. The k^{th} stage of the solution provides the sample size for the k^{th} stratum. The compromise allocation thus obtained is optimum in the sense that it minimizes the weighted sum of the sampling variances of the estimates of the population means of various characteristics.

This chapter is based on my joint research paper entitled "On compromise allocation in multivariate stratified sampling" submitted for publication in the Naval Research Logistics (vide their manuscript number 3280).

4.2 THE PROBLEM

When the total amount available for a multivariate stratified survey is prefixed, a compromise allocation may be that which minimizes the weighted sum of the sampling variances of the estimates of various characteristics with in the available budget. It is assumed that the characteristics are mutually independent hence the covariances are zero. Let the given population be divided into L strata and there be p independent characteristics defined on every population unit. Hereinafter all the notations used (except c_h) are as defined in section 4.1 except for the additional suffix 'j' which indicates that the quantity corresponds to the j^{th} characteristic; j=1,2,...,p.

If the population means of various characteristics are of interest, it may be a reasonable criterion for obtaining the compromise allocation to minimize the weighted sum

$$\sum_{j=1}^{p} a_j V(\bar{y}_{jst})$$

where $$V(\bar{y}_{jst}) = \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{N_h} \; ; \; j=1,2,\ldots,p \quad (4.5)$$

is the sampling variance of the stratified sample mean $\bar{y}_{jst}$. Where

$\bar{y}_{jst} = \sum_{h=1}^{L} W_h \bar{y}_{jh}$ is the estimate of $\bar{Y}_j$.,

$\bar{Y}_j = \sum_{h=1}^{L} W_h \bar{Y}_{jh}$ is the over all population mean for j^{th} characteristic,

$S_{jh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{jhi} - \bar{Y}_j)^2$ is the stratum variance for the j^{th} characteristic in the h^{th} stratum

and $a_j > 0$ are weights assigned to various characteristics according to some measure of their importance.

It is conjectured that weights a_j; $j=1,2,\ldots,p$ should be proportional to the sum of the stratum variances for the j^{th} characteristic, that is

$$a_j \; \alpha \; \sum_{h=1}^{L} S_{jh}^2 ; \; j=1,2,\ldots,p.$$

Letting $\sum_{j=1}^{p} a_j = 1$, the above conjecture leads to:

$$a_j = \frac{\sum_{h=1}^{L} S_{jh}^2}{\sum_{j=1}^{p} \sum_{h=1}^{L} S_{jh}^2} \; ; \; j=1,2,\ldots,p \qquad (4.6)$$

With a linear cost function "$c_o + \sum c_h n_h$" the problem of finding the compromise allocation for a fixed cost C may be given as the following NLMPP:

$$\text{Minimize} \sum_{j=1}^{p} a_j V(\bar{y}_{jst}) = \sum_{j=1}^{p} a_j \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{n_h} - \sum_{h=1}^{p} a_j \sum_{h=1}^{L} \frac{W_h^2 S_{jh}^2}{N_h}$$

$$\text{subject to} \quad c_o + \sum_{h=1}^{L} c_h n_h \leq C \qquad (4.7)$$

$$\text{and} \quad 2 \leq n_h \leq N_h; \; h=1,2,\ldots,L.$$

Where n_h; $h=1,2,\ldots,L$ is the required compromise allocation assumed to be continuous over the interval $[2,N_h]$; $h=1,2,\ldots,L$,

$$c_h = \sum_{j=1}^{p} c_{jh}$$

is the cost of measuring all the p characteristics on a sampled unit form the h^{th} stratum and c_{jh} is the per unit cost of measuring the j^{th} characteristic in the h^{th} stratum.

Note that the above definition of c_h is different from that used in (4.3). Hereinafter by c_h we mean the c_h as defined above.

The terms under the second summation in the objective function of the NLMPP (4.7) are constants with respect to n_h and therefore may be ignored for the purpose of minimization. Ignoring the terms independent of n_h, interchanging the order of the summation and letting

$$B_h = W_h^2 \sum_{j=1}^{p} a_j S_{jh}^2;\ j=1,2,\ldots,L \qquad (4.8)$$

the objective function of the NLMPP (4.7) may be restated as:

$$\text{Minimize } Z(n_1,n_2,\ldots,n_L) = \sum_{h=1}^{L} \frac{B_h}{n_h} \qquad (4.9)$$

$$\text{subject to } \sum_{h=1}^{L} c_h n_h \leq C_o \qquad (4.10)$$

$$\text{and } 2 \leq n_h \leq N_h;\ h=1,2,\ldots,L \qquad (4.11)$$

where $C_o=C-c_o$.

The restrictions (4.11) are imposed to avoid over sampling, that is the situation where $n_h \geq N_h$ and to estimate the stratum variances S_{jh}^2 for which at least two units are to be selected from each stratum.

A careful study of the NLMPP (4.9) - (4.11) reveals its following special features:

(i) Since $B_h>0$; $h=1,2,\ldots,L$ the objective function Z is strictly convex.

(ii) The feasible region given by

$$F=\{(n_1,n_2,\ldots,n_L) \mid \sum_{h=1}^{L} c_h n_h \leq C_o \quad \text{and} \quad 2\leq n_h \leq N_h;\ h=1,2,\ldots,L\}$$ is a *bounded convex region in n-dimensional Euclidean space* E_n.

(iii) The necessary and sufficient condition for F to be nonempty and hence for the existence of an optimal solution is $2\sum_{h=1}^{L} c_h \leq C_o$.

(iv) The optimum solution of NLMPP (4.9)-(4.11), if it exists, will be unique and will be a point on the boundary $\sum_{h=1}^{L} c_h n_h = C_o$ of the feasible region F. In other words the constraint (4.10) will be active at the optimal point. This is due to the convexity of the objective function and the linearity of the constraints.

4.3 THE DYNAMIC PROGRAMMING APPROACH

If the restrictions in (4.11) are ignored NLMPP (4.9)-(4.11) reduces to

"Minimize $$\sum_{h=1}^{L} \frac{B_h}{n_h}$$

subject to $$\sum_{h=1}^{L} c_h n_h = C_o$$"

Lagrange multipliers techniques may be used to find the optimum n_h; h=1,2,...,L. Define $\phi(n,\lambda) = \sum_{h=1}^{L} \frac{B_h}{n_h} + \lambda\left(\sum_{h=1}^{L} c_h n_h - C_o\right)$. Differentiating ϕ with respect to n_h; h=1,2,...,L and λ and equating to zero we get the following L+1 simultaneous equations.

$$\frac{\partial \phi}{\partial n_h} = -\frac{B_h}{n_h^2} + \lambda c_h = 0;\ h=1,2,\ldots,L$$

and

$$\frac{\partial \phi}{\partial \lambda} = \sum_{h=1}^{L} c_h n_h - C_o = 0.$$

These equations together give the optimum value of n_h i.e. n_h^*; h=1,2,...,L as:

$$n_h^* = \frac{C_o\sqrt{B_h/c_h}}{\sum_{h=1}^{L}\sqrt{B_h c_h}};\ h=1,2,\ldots,L \qquad (4.12)$$

If the above values of n_h^* satisfy (4.11) also the NLMPP (4.9)-(4.11) is solved and (4.12) will give the required compromise allocation. In case some or all of the n_h^* given by (4.12) violates (4.11), the Lagrange multipliers technique fails and some other constrained optimization technique is to be used. In the following a procedure to obtain the compromise allocation using the dynamic programming technique in case the Lagrange multipliers technique fails is developed. For problems whose

solutions may be obtained by using Lagrange multipliers technique the dynamic programming approach gives an identical solution.

The objective function as well as the constraint of the NLMPP (4.9)-(4.11) are sum of independent functions of n_h; h=1,2,...,L. The NLMPP which is an L-stage decision problem can be decomposed into L single-stage single variable decision problems.

If Z^* denote the optimal value of the objective function (4.9) under the constraints (4.10) and (4.11) then

$$Z^* = \min_{n_1, n_2, \ldots, n_L} \left[\sum_{h=1}^{L} \frac{B_h}{n_h} \right] = f'(C_L) \text{ (say)}, \qquad (4.13)$$

where the minimization is carried out over the set of the feasible solutions

$$F=\{(n_1, n_2, \ldots, n_L) \mid \sum_{h=1}^{L} c_h n_h \le C_o \text{ and } 2 \le n_h \le N_h;\ h=1,2,\ldots,L\}$$

Note that $C'_L = C_o$.

Again let n_L be any feasible value of the L^{th} decision variable. Keeping n_L fixed we then compute

$$\min_{n_1, n_2, \ldots, n_{L-1}} \left[\sum_{h=1}^{L} \frac{B_h}{n_h} \right] = \frac{B_L}{n_L} + \min_{n_1, n_2, \ldots, n_{L-1}} \left[\sum_{h=1}^{L-1} \frac{B_h}{n_h} \right] \qquad (4.14)$$

Then $f(C'_L)$ will be the smallest of all the RHS values given by (4.14) for all feasible n_L.

Now

$$f(C'_{L-1}) = \min_{n_1, n_2, \ldots, n_{L-1}} \left[\sum_{h=1}^{L-1} \frac{B_h}{n_h} \right] \tag{4.15}$$

where the minimization is carried out over the set of feasible solutions

$$\{(n_1, n_2, \ldots, n_L) \mid \sum_{h=1}^{L-1} c_h n_h \leq C_o - c_L n_L \text{ and } 2 \leq n_h \leq N_h;\ h=1,2,\ldots,L-1\}$$

and C_{L-1} denotes the available budget for (L-1) strata. Obviously

$$C'_{L-1} = C'_L - c_L n_L .$$

Combining (4.13), (4.14) and (4.15)

$$f(C'_L) = \min_{n_L} \left[\frac{B_L}{n_L} + f(C'_{L-1}) \right], \tag{4.16}$$

where n_L takes on values

$$2 \leq n_L \leq \min(C''_L, N_L) \tag{4.17}$$

and C''_L = maximum possible size of the sample that can be drawn from the L^{th} strata within the available budget C'_L i.e.

$$C''_L = \frac{C'_L - 2\sum_{h=1}^{L-1} c_h}{c_L} \tag{4.18}$$

To evaluate $f(C_L)$ the RHS of (4.16) is to be minimized with respect to a single variable n_L given by (4.17) provided $f(C'_{L-1})$ is known.

To compute $f(C'_{L-1})$ one can proceed just as above and get

$$f(C'_{L-1}) = \min_{n_{L-1}} \left[\frac{B_{L-1}}{n_{L-1}} + f(C'_{L-2})\right], \tag{4.19}$$

where

$$f(C'_{L-2}) = \min_{n_1, n_2, \ldots, n_{L-2}} \left[\sum_{h=1}^{L-2} \frac{B_h}{n_h}\right], \tag{4.20}$$

and the minimization is carried out over the set of feasible solutions

$$\{(n_1, n_2, \ldots, n_{L-2}) \mid \sum_{h=1}^{L-2} c_h n_h \leq C_o - (c_{L-1} n_{L-1} + c_L n_L) \text{ and } 2 \leq n_h \leq N_h; \; h = 1, 2, \ldots, L-2\}$$

This procedure is continued until we evaluate

$$f(C'_1) = \min_{n_1} \frac{B_1}{n_1} \tag{4.21}$$

In actual practice we first evaluate $f(C'_1)$ then $f(C'_2)$ and so on and finally $f(C'_L)$ or Z^*.

At the k^{th} stage of solution we have to find n_k^* for which

$$f(C_k') = \min_{n_k} \left[\frac{B_k}{n_k} + f(C_{k-1}') \right], \tag{4.22}$$

where

$$f(C_k') = \left[\min \sum_{h=1}^{k} \frac{B_h}{n_h} \mid \sum_{h=1}^{k} c_h n_h \leq C_k' \text{ and } 2 \leq n_h \leq N_h;\ h=1,2,\ldots,k \right] \tag{4.23}$$

for all C_k' satisfying

$$2 \sum_{h=1}^{k} c_h \leq C_k' \leq C_o - 2 \sum_{h=k+1}^{L} c_h . \tag{4.24}$$

C_k' denote the cost available for measuring all the units selected in the sample from first k strata.

The RHS of (4.22) is minimized over n_k given by

$$2 \leq n_k \leq \min(C_k'', N_k) \tag{4.25}$$

where C_k''=maximum possible size of the sample which could be drawn from the k^{th}strata within the available cost C_k' for first k strata i.e.,

$$C_k'' = \frac{C_k' - 2 \sum_{h=1}^{k-1} c_h}{c_k} \tag{4.26}$$

Initially we set $f(C'_o)=0$.

From the above discussion it is clear that solving NLMPP (4.9)-(4.11) is equivalent to find $f(C'_L)$. Using the recurrence formula (4.22) $f(C'_L)$ is found recursively. From $f(C')$, n_L^* is computed; similarly n_{L-1}^* is computed from $f(C'_{L-1})$; and so on until finally n_1^* is obtained.

Assuming n_h as continuous variables, at the k^{th} stage; $k=1,2,\ldots,L$, differential calculus may be used to minimize $\left[\frac{B_k}{n_k} + f(C'_{k-1})\right]$, provided the n_k^* so obtained remain feasible, otherwise some search method is to be used.

The following two examples illustrate the computational details of the solution procedure developed by the authors using the dynamic programming technique. The data of the first example is realistic and is due to Jessen (1942) as reported in Sukhatme et al (1984) except for the costs c_h and C which are assumed by the authors. The measurement cost is assumed to be variable while in Jessen (1942) they are constants. It is also assumed that the overhead cost $c_o=500$ units while the total amount available for survey C=4500.

4.4 NUMERICAL EXAMPLES

Example 1: In a stratified population with five strata the population means of three independent characteristics are to be estimated. The values of N_h, W_h, the estimated stratum variances

s_{1h}^2, s_{2h}^2, and s_{3h}^2 and the cost c_h are given in the Table 4.1.

The value of C, the total cost available for the survey, is assumed to be 4500 units while the overhead cost c_o is 500 units. Thus the total cost available for measurements C_o=4500-500=4000 units.

Table 4.1

Data for 5 strata and 3 characteristics

h	N_h	W_h	s_{1h}^2	s_{2h}^2	S_{3h}^2	c_h
1	39552	0.197	12	56	41.3	2
2	38347	0.191	80	2132	23.1	3
3	43969	0.219	1113	565	10.9	4
4	36942	0.184	84	355	11.5	5
5	41760	0.208	247	68	38.8	6

From Table 4.1

$$\sum_{h=1}^{5} s_{1h}^2 = 1536,$$

$$\sum_{h=1}^{5} s_{2h}^2 = 3176,$$

$$\sum_{h=1}^{5} s_{3h}^2 = 125.6$$

and

$$\sum_{h=1}^{5} \sum_{j=1}^{3} s_{jh}^2 = 4837.6$$

Replacing s_{jh}^2 by their sample estimates s_{jh}^2 in (4.6) the weights a_j; j=1,2,3 are worked out as:

$$a_1 = \frac{1536}{4837.6} = 0.3175,$$

$$a_2 = \frac{3176}{4837.6} = 0.6565$$

and $$a_3 = \frac{125.6}{4837.6} = 0.0260.$$

The values of the coefficients B_h; h=1,2,...,5 given by (4.8) approximated to four places of decimal are worked out in Table 4.2

Table 4.2

Calculation of B_h for 5 strata

h	W_h	W_h^2	j=1 $a_1 S_{1h}^2$	j=2 $a_2 S_{2h}^2$	j=3 $a_3 S_{3h}^2$	$\sum_{j=1}^{3} a_j S_{jh}^2$	$B_h = W_h^2 \sum a_j S_{jh}^2$
1	0.197	0.038809	3.81	36.764	1.0738	41.6478	1.6163
2	0.191	0.036481	25.4	1399.658	0.6006	1425.6586	52.0095
3	0.219	0.047961	353.3775	370.9225	0.2834	724.5834	34.7517
4	0.184	0.033856	26.67	233.0575	0.299	260.0265	8.8035
5	0.208	0.043264	78.4225	44.642	1.0088	124.0733	5.3679

The NLMPP (4.9)-(4.11) for the numerical values given in Table 4.1 and the values of B_h; h=1,2,...,5 given in the Table 4.2 is formulated as:

$$\text{Minimize } Z = \frac{1.6163}{n_1} + \frac{52.0095}{n_2} + \frac{34.7517}{n_3} + \frac{8.8035}{n_4} + \frac{5.3679}{n_5} \quad (4.27)$$

$$\text{subject to } 2n_1+3n_2+4n_3+5n_4+6n_5 \leq 4000 \quad (4.28)$$

$$2 \leq n_1 \leq 39522,$$

$$2 \leq n_2 \leq 38347,$$

$$2 \leq n_3 \leq 43967, \quad (4.29)$$

$$2 \leq n_4 \leq 36942,$$

$$2 \leq n_5 \leq 41760.$$

Ignoring (4.29) and taking equality in (4.28), the optimum sample sizes n_h^*; h=1,2,...,5 (rounded off to the nearest integer values) using (4.12) are worked out as:

Table 4.3

Calculation of n_h^* using formula (4.12)

h	B_h	c_h	$\sqrt{B_h/c_h}$	$\sqrt{B_h c_h}$	n_h^* (rounded off)
1	1.6163	2	0.8990	1.7979	94
2	52.0095	3	4.1637	12.4911	434
3	34.7517	4	2.9475	11.7901	307
4	8.8035	5	1.3269	6.6346	138
5	5.3679	6	0.9459	5.6752	98

$$\sum_{L} \sqrt{B_h c_h} = 38.3889$$

These values of n_h^* satisfy (4.29) also, hence they will solve the NLMPP (4.27)-(4.29) completely. The optimal value Z^* of the

objective function Z, is $Z^{*}=0.3684$.

For the sake of illustration, in the following, the dynamic programming approach to the NLMPP (4.27)-(4.29) is given.

As defined earlier C'_h; h=1,2,...,5 and their limits are:

$$C'_5 = 2n_1+3n_2+4n_3+5n_4+6n_5 = 4000$$

$$C'_4 = C'_5-6n_5;\ 28\le C'_4\le 3988$$

$$C'_3 = C'_4-5n_4;\ 18\le C'_3\le 3978$$

$$C'_2 = C'_3-4n_3;\ 10\le C'_2\le 3970$$

$$C'_1 = C'_2-3n_2;\ 4\le C'_1\le 3964.$$

The values of N_h; h=1,2,...,5 are sufficiently large to assume that $\min(C''_k,N_k)=C''_k$, where C''_k is as defined by (4.26).

For the first stage of solution

$$f(C'_1) = \min_{2\le n_1\le C''_1}\left[\frac{1.6163}{n_1} = f(C'_o)\right]$$

$$= \min_{2\le n_1\le C''_1}\left[\frac{1.6163}{n_1}\right], \text{ because } f(C'_o)=0$$

$$\Rightarrow \quad f(C'_1) = \frac{3.3226}{C''_1}, \text{ at } n_1^{*} = \frac{C'_1}{2} \qquad (4.30)$$

For the second stage of solution

$$f(C_2') = \min_{2 \le n_2 \le C_2''} \left[\frac{52.0095}{n_2} + f(C_1') \right]$$

$$= \min_{2 \le n_2 \le (C_2' - 2c_1)/c_2} \left[\frac{52.0095}{n_2} + \frac{3.232}{C_1'} \right]$$

$$= \min_{2 \le n_2 \le (C_2' - 4)/3} \left[\frac{52.0095}{n_2} + \frac{3.2326}{C_2' - 3n_2} \right]$$

$$\Rightarrow \quad f(C_2') = \frac{204.1778109}{C_2'}, \quad \text{at } n_2^* = 0.291391205\ C_2' \tag{4.31}$$

The expression (4.31) is obtained by using differential calculus for minimizing the quantity inside [ ] with respect to n_2 for values of C_2' satisfying $10 \le C_2' \le 3970$.

Similarly for the third and fourth stages of solution we get

$$f(C_3') = \frac{680.1243926}{C_3'}, \quad \text{at } n_3^* = 0.113022225\ C_3' \tag{4.32}$$

and

$$f(C_4') = \frac{1070.190302}{C_4'}, \quad \text{at } n_4^* = 0.040561329\ C_4' \tag{4.33}$$

respectively.

For the fifth and final stage of solution $f(C_5')$ is obtained

$$f(C_5') = 0.368427286 \quad \text{at } n_5^* = 98.55545903 \tag{4.34}$$

Using the values of n_5^* given by (4.34) we get

$$C_4' = C_5' - 6n_5^* = 4000 - 6 \times 98.55545903 = 3408.667246$$

Substituting this value of C_4' in (4.33) we get

$$n_4^* = 0.040561329 \times 3408.667246 = 138.2600736$$

Proceeding in this manner we obtain

$$n_3^* = 307.1228507,$$

$$n_2^* = 433.8452188$$

and $$n_1^* = 93.6699094.$$

Rounding off to their nearest integer values the optimum compromise allocations are obtained as:

$$n_1^* = 94,\ n_2^* = 434,\ n_3^* = 307,\ n_4^* = 138 \text{ and } n_5^* = 98.$$

Which are same as calculated in the Table 4.3 by using formula (4.12).

Example 2: In a stratified population with three strata and two independent characteristics the values of N_h, W_h, S_{1h}, S_{2h} and c_h are as given in the Table 4.4

Table 4.4

Data for 3 strata and 2 characteristics

h	N_h	W_h	S_{1h}	S_{2h}	C_h
1	18	0.30	2	3	3
2	27	0.45	4	1	4
3	15	0.25	20	35	5

Assuming both the characteristics equally important, that is $a_1=a_2=1$, the problem of finding a compromise allocation for a total fixed budget C=125 units including an overhead cost $c_o=25$ units, may be expressed as

$$\text{Minimize} \quad Z = \frac{1.1700}{n_1} + \frac{3.4425}{n_2} + \frac{101.5625}{n_3} \tag{4.35}$$

$$\text{subject to} \quad 3n_1+4n_2+5n_3 \leq 100 \tag{4.36}$$

and

$$2 \leq n_1 \leq 18$$

$$2 \leq n_2 \leq 27 \tag{4.37}$$

$$2 \leq n_3 \leq 15.$$

Table 4.5

h	$\sum_{j=1}^{p} a_j S_{jh}^2$	$B_h=W_h^2 \sum_{j=1}^{p} a_j S_{jh}^2$	$\sqrt{B_h c_h}$	$\sqrt{B_h/c_h}$	n_h^*
1	13	1.1700	1.8735	0.6245	2.2209
2	17	3.4425	3.7108	0.9277	3.2992
3	1625	101.5625	22.5347	4.5069	16.0279

$$\sum \sqrt{B_h c_h} = 28.1190$$

The rounded off solution given by the last column of the Table 4.5 is:

$$n_1^*=2,\ n_2^*=3,\ \text{and}\ n_3^*=16.$$

This solution is infeasible because it violates the restriction $2\le n_3^*\le 15$ in (4.37).

In the above situation dynamic programming may be used as an alternative.

We have C_k'; k=1,2,3 and their limits as:

$$C_3' = 3n_1+4n_2+5n_3 = 100,$$

$$C_2' = C_3'-5n_3;\ 14\le C_2'\le 90,$$

and $$C_1' = C_2'-4n_2;\ 6\le C_1'\le 54.$$

For the first stage of solution

$$f(C_1') = \min_{2\le n_1\le \min(C_1'',N_1)} \left[\frac{1.17}{n_1} + f(C_o')\right]$$

$$= \min_{2\le n_1\le \min\left(\frac{C_1'}{n_1},18\right)} \left[\frac{1.17}{n_1}\right], \text{ because } f(C_o')=0$$

$$= \min_{2\le n_1\le C_1'/3} \left[\frac{1.17}{n_1}\right]$$

(Using limits of C_1' it can be seen that $\min\left(\frac{C_1'}{3},18\right) = \frac{C_1'}{3}$)

$$\Rightarrow f(C_1') = \frac{3.51}{C_1'} \text{ , at } n_1^* = 0.3333\ C_1' \tag{4.38}$$

For the second stage of solution

$$f(C_2') = \min_{2 \le n_2 \le \min(C_2'', N_2)} \left[\frac{3.4425}{n_2} + \frac{3.512}{C_1'} \right]$$

$$= \min_{2 \le n_2 \le \left(\frac{C_2'-6}{4}\right)} \left[\frac{3.4425}{n_2} + \frac{3.51}{C_2'-4n_2} \right]$$

$$f(C_2') = \frac{31.1843}{C_2'} \text{ , at } n_2^* = 0.1661\ C_2' \tag{4.39}$$

For the third and final stage of solution

$$f(C_3') = \min_{2 \le n_3 \le \min(C_3'', N_3)} \left[\frac{101.5625}{n_2} + \frac{31.1843}{C_2'} \right]$$

$$= \min_{2 \le n_3 \le \min\left(\frac{C_3'-14}{5}, 15\right)} \left[\frac{101.5625}{n_3} + \frac{31.1843}{C_3'-n_3} \right]$$

$$= \min_{2 \le n_3 \le \min\left(\frac{100-14}{5}, 15\right)} \left[\frac{101.5625}{n_3} + \frac{31.1843}{100-n_3} \right]$$

$$= \min_{2 \le n_3 \le 15} \left[\frac{101.5625}{n_3} + \frac{31.1843}{100-n_3} \right]$$

$$\Rightarrow f(C_3') = 8.0182, \text{ at } n_2^* = 15 \tag{4.40}$$

Now $C_2'=C_3'-c_3n_3^* = 100-5x15 = 25.$

Thus by (4.40)

$$n_2^* = 0.1661x25 = 4.1525.$$

Again $C_1'=C_2'-c_2n_2^* = 25-4x4.1525 = 8.39$

By (4.38)

$$n_1^* = 0.3333x8.39 = 2.7966$$

Rounding off to the nearest integer value of the optimum compromise allocation is given as:

$$n_1^* = 3,\ n_2^* = 4 \text{ and } n_3^* = 15, \quad \text{with } Z^*=8.0215.$$

4.5 DISCUSSION

The NLMPP (4.7) provides a general formulation of the problem of obtaining a compromise allocation in multivariate stratified random sampling with p independent characteristics. The following situations are its particular cases.

(1) With $a_j=1;\ j=1,2,\ldots,p$, the objective of NLMPP (4.7) will become

$$\text{"Minimize} \quad \sum_{j=1}^{p} V(\bar{y}_{jst})\text{"}$$

which is equivalent to minimize the trace of the variance covariance matrix of $\bar{y}_{jst};\ j=1,2,\ldots,p$ (which is a diagonal matrix) because the characteristics are independent) for a fixed

budget.

(2) With $c_h=1$; $h=1,2,\ldots,L$ and $C_o=C-c_o=n$ (the total sample size) and taking equality the constraint would become

$$" \sum_{h=1}^{L} n_h = n"$$

in which case the compromise allocation would be for a fixed total sample size.

(3) Let the loss function $l(z_j)$ due to an error $z_j=(\bar{y}_{jst}-\bar{Y}_j)$ in the estimate $\bar{y}_{jst}$ of $\bar{Y}_j$ be a sample quadratic function of z_j, that is

$$l(z_j) = b_j Z_j^2$$

$$= b_j(\bar{y}_{jst}-\bar{Y}_j)^2$$

where $b_j>0$; $j=1,2,\ldots,p$ are known constants.

The expected loss L_j; $j=1,2,\ldots,p$ in this case would be

$$L_j = E[b_j(\bar{y}_{jst}-\bar{Y}_j)^2]$$

$$= b_j E(\bar{y}_{jst}-\bar{Y}_j)^2$$

$$= b_j V(\bar{y}_{jst}).$$

The objective of the NLMPP (4.7) would be to minimize the

total expected loss for a fixed budget.

(4) For practical implementation of any allocation we need integer values of the sample sizes from various strata. The integer values may be obtained by rounding off the noninteger values. Often these integer values of the sample sizes become infeasible or nonoptimal. In such situations integer restrictions may also be imposed on the variables n_h; h=1,2,...,L in the NLMPP (4.7) and we have to solve an All Integer Nonlinear Programming Problem (AINLPP). The procedure developed in this chapter has this added advantage that it could be modified to obtain the integer optimum compromise allocations.

In the last Chapter of this thesis an integer optimum solution to the NLMPP discussed in this chapter is worked out using dynamic programming technique.

CHAPTER-V

DETERMINING THE OPTIMUM ALLOCATION IN MULTIVARIATE STRATIFIED SAMPLING : AN INTEGER SOLUTION

5.1 INTRODUCTION

As discussed in Chapter-IV for practical application of any allocation integer values of the sample sizes are required. This could be obtained by simply rounding off noninteger sample sizes to their nearest integral values. When the sample sizes are large enough and (or) the measurement costs in various strata are not too high, the rounded off sample allocations may work well. However in situations other than described above the rounded off sample allocations may become infeasible and nonoptimal. This means that the rounded off values may violate the cost constraint and (or) there may exist other sets of integer sample allocations with a lesser value of the objective function defined elsewhere in this manuscript.

In this chapter the problem of obtaining a compromise allocation in multivariate stratified random sampling as formulated in Chapter-IV is considered as an All Integer Nonlinear Programming Problem (AINLPP). This AINLPP is treated as a multistage decision problem and a solution procedure is developed using the dynamic programming technique in which the k^{th} stage of

the solution provides the required size for the k^{th} stratum.

This chapter is based on my research paper entitled "An optimal multivariate stratified sampling design using dynamic programming" presented in the 3rd International Triennial Calcutta Symposium held in December, 1997. The paper is also due to appear in the proceedings of the above Symposium to be published by Wiley Publications.

5.2 THE PROBLEM

With integer restrictions on n_h the NLMPP (4.9)-(4.11) formulated in Chapter-IV of this thesis will become the AINLPP

$$\text{Minimize} \quad Z(n_1, n_2, \ldots, n_L) = \sum_{h=1}^{L} \frac{W_h^2 A_h^2}{n_h} \tag{5.1}$$

$$\text{subject to} \quad \sum_{h=1}^{L} c_h n_h \leq C_o \,, \tag{5.2}$$

$$2 \leq n_h \leq N_h \tag{5.3}$$

$$\text{and } n_h \text{ integer; } h=1,2,\ldots,L. \tag{5.4}$$

where

$$A_h^2 = \sum_{j=1}^{p} a_j S_{jh}^2; \; h=1,2,\ldots,L. \tag{5.5}$$

Note that the objective functions (4.9) and (5.1) are same

because B_h used in (4.9) is nothing but $W_h^2 A_h^2$. This has been done just to differentiate between the two objective functions only.

5.3 THE SOLUTION

The objective function and the constraints of the AINLPP (5.1)-(5.4) are separable functions of n_h; h=1,2,...,L.

Due to the separability of the functions and the nature of the problem of allocation the dynamic programming technique may be used to solve the AINLPP (5.1)-(5.4) (see Hadley (1964)).

In the following a solution procedure for solving AINLPP (5.1)-(5.4) using dynamic programming technique is presented.

Consider the subproblem called the k^{th} subproblem involving the first k(<L) strata as:

$$\text{Minimize} \quad \sum_{h=1}^{k} \frac{W_h^2 A_h^2}{n_h} \tag{5.6}$$

$$\text{subject to} \quad \sum_{h=1}^{k} c_h n_h \leq C_k, \tag{5.7}$$

$$2 \leq n_h \leq N_h \tag{5.8}$$

$$\text{and} \quad n_h \text{ integer; } h=1,2,\ldots,k. \tag{5.9}$$

Where C_k<C is the available amount for the first k strata. Note that C_k=C if k=L.

Let $f(k,C_k)$ be the minimum value of the objective function of the problem (5.6)-(5.9), then

$$f(k,C_k) = \left\{\min \sum_{h=1}^{k} \frac{W_h^2 A_h^2}{n_h} \middle| \sum_{h=1}^{k} c_h n_h \le C_k,\ 2 \le n_h \le N_h \text{ and } n_h \text{ are integers } h=1,2,\ldots,k\right\} \quad (5.10)$$

with this definition of $f(k,C_k)$ the AINLPP (5.1)-(5.4) is equivalent to find $f(L,C)$, which can be obtained by finding $f(k,C_k)$ recursively for $k=1,2,\ldots,L$ and for all feasible C_k, that is, $2\sum_{h=1}^{k} c_h \le C_k \le C$.

We can express (5.10) as:

$$f(k,C_k) = \left\{\min\left(\frac{W_k^2 A_k^2}{n_k} + \sum_{h=1}^{k-1} \frac{W_h^2 A_h^2}{n_h}\right) \middle| \sum_{h=1}^{k-1} c_h n_h \le C_k - c_k n_k,\ 2 \le n_h \le N_h \text{ and } n_h \text{ are integer; } h=1,2,\ldots,k\right\}$$

For a fixed integer value of n_k, $2 \le n_k \le \min\left(\left[\frac{C_k}{c_k}\right], N_k\right)$, where $\left[\frac{C_k}{c_k}\right]$ is the largest integer $\le \frac{C_k}{c_k}$, $f(k,C_k)$ is given by

$$f(k,C_k) = \frac{W_k^2 A_k^2}{n_k} + \left\{\min \sum_{h=1}^{k-1} \frac{W_h^2 A_h^2}{n_h} \middle| \sum_{h=1}^{k-1} c_h n_h \le C_k - c_k n_k,\ 2 \le n_h \le N_h \text{ and } n_h \text{ are integer; } h=1,2,\ldots,k\right\} \quad (5.11)$$

By the definition (5.10) the quantity inside $\{\ \}$ in (5.11) is $f(k-1,C_{k-1})$, where $C_{k-1}=C_k-c_k n_k$. Thus the required recurrence relation is

$$f(k,C_k) = \min_{n_k \in I_k} \left[\frac{W_k^2 A_k^2}{n_k} + f(k-1,C_{k-1})\right] \qquad (5.12)$$

$$\text{where } I_k = \left\{ n_k \,\middle|\, 2 \le n_k \le \min\left(\left[\frac{C_k}{c_k}\right], N_k\right),\ n_k \text{ integer} \right\} \qquad (5.13)$$

At the final stage of the solution i.e. at k=L, f(L,C) is obtained by solving (5.12) recursively for all C_k. From f(L,C) the optimum value n_L^* of n_L is obtained, from $f(L-1,C_{L-1})$ the optimum value of n_{L-1}^* of n_{L-1} is obtained and so on until finally we obtain the optimum value n_1^* of n_1.

We also define

$$f(k,C_k) = 0 \quad \text{for } k=0 \qquad (5.14)$$

and $$f(k,C_k) = \infty \quad \text{if } C_k < 2\sum_{h=1}^{k} c_h \text{ or } n_k > N_k \quad k=1,2,\ldots,L. \qquad (5.15)$$

It is to be noted that (5.15) takes care of the restrictions $2 \le n_h \le N_h$; h=1,2,...,L of the AINLPP (5.1)-(5.4).

5.4 A NUMERICAL EXAMPLE

The following numerical example demonstrates the use of the solution procedure. The data used in this example is from a

stratified random sample survey conducted in Varanasi district of Uttar Pradesh (U.P), India to study the distribution of manurial resources among different crops and cultural practices (see Sukhamte et al (1984)). Relevant data with respect to the two characteristics "area under rice" and "total cultivated area" are given in Table 5.1. The total number of villages in the district was 4190.

Table 5.1

Data for four strata and two characteristics

h	N_h	W_h	S_{h1}^2	S_{h2}^2
1	1419	.3387	4817.72	130121.15
2	619	.1477	6251.26	7613.52
3	1253	.2990	3066.16	1456.40
4	899	.2146	56207.25	66977.72

In addition to the above information to demonstrate the procedure the following are also assumed. The per unit cost of measurement c_h in various strata are assumed as $c_1=3$, $c_2=4$, $c_3=5$ and $c_4=6$ units. The total amount available for the survey C_o is assumed as 2400 units including an expected overhead cost $c_o=400$ units. The total amount available for measurements is thus C=2400-400=2000 units.

From the Table 5.1 $$\sum_{h=1}^{4} S_{h1}^2 = 70342.39$$

and $$\sum_{h=1}^{4} S_{h2}^2 = 206168.79$$

Using formula (4.6) that is

$$a_j = \frac{\sum_{h=1}^{L} S_{jh}^2}{\sum_{j=1}^{p}\sum_{h=1}^{L} S_{jh}^2} \; ; \; j=1,2,\ldots,p$$

of Chapter-IV the wiights a_j; j=1,2 are obtained as

$$a_1 = \frac{70342.39}{70342.39+206168.79} \cong 0.25$$

$$a_2 = \frac{206168.79}{70342.39+206168.79} \cong 0.75$$

Using (5.5) A_h^2; h=1,2,3 and 4 are worked out as:

$$A_1^2 = 98795.30222, \qquad A_2^2 = 7272.951299,$$

$$A_3^2 = 1858.842864, \qquad A_4^2 = 64285.11773.$$

Substituting the above values of A_h^2, W_h^2, c_h, N_h; h=1,2,3 & 4 and C in (2.5) we get the following AINLPP:

$$\text{Minimize } Z(n_1,n_2,n_3,n_4) = \frac{11333.5688}{n_1} + \frac{158.6615}{n_2} + \frac{166.1824}{n_3} + \frac{2960.5328}{n_4} \qquad (5.16)$$

subject to $3n_1+4n_2+5n_3+6n_4 \leq 2000,$ (5.17)

$$2 \leq n_1 \leq 1419,$$

$$2 \leq n_2 \leq 619, \quad (5.18)$$

$$2 \leq n_3 \leq 1253,$$

$$2 \leq n_4 \leq 899,$$

and n_h integer; h=1,2,3,4. (5.19)

The computer program (in 'C' language) of the procedure developed in section 5.3 for solving the AINLPP (5.16)-(5.19) is as given below.

```
#include<stdio.h>
#include<string.h>
main()
{
 int p,b,nk,c,n[6][2010],j,k_max=4,c1,k,c_max=2000,m,n0;
 float f[5][2010],pvf,al,min;
 long int NK[5]={1,1419,619,1253,899};
 float wk[5]={1,0.3387,0.1477,0.2990,0.2146};
 float ak[5]={1,314.3172,85.2816,43.1143,253.5451};
 int ck[5]={1,3,4,5,6};
  f[1][0]=9999999.0;
  f[2][0]=9999999.0;
  f[3][0]=9999999.0;
  for(k=1;k<=k_max;k++)
  {
  c1=0;
  for(j=1;j<=k;j++)
  c1=c1+ck[j];
  c1=2*c1;
  for(c=1;c<=c_max;c++)
  {
   p=c/ck[k];
   if(c < c1)
   {
    f[k][c]=9999999;
    n[k][c]=n;
```

```
    }
    else
    {
     min=9999999.0;
     if(p>NK[k])
      p=NK[k];
     for(nk=2;nk<=p;nk++)
      {
       if(k==1)
        f[k-1][c-ck[k]*nk]=0.0;
       else
       pvf=f[k-1][c-ck[k]*nk];
       if(pvf<=9999999.0)
        f[k][c]=9999999.0;
       else
        f[k][c]=((wk[k]*wk[k])*(ak[k]*ak[k]))/nk+f[k-1][c-ck[k]*nk];
        if(f[k][c] < min)
        {
         min=f[k][c];
         al=f[k-1][c-ck[k]*nk];
         n[k][c]=nk;
        }
      } /*loop for nk*/
      f[k][c]=min;
     } /* loop for else*/
   } /* loop for c*/
   } /* loop for k*/
 m=c_max;
 for(k=k_max;k>0;k--)
 {
 printf("\t\tThe result n[%d][%d]=%d\n",k,m,n[k][m]);
 n0=ck[k]*n[k][m];
 m=m-n0;
 }
 }
```

Execution of the above program gives the following results.

$$n_1^*=331,\ n_2^*=33,\ n_3^*=31,\ n_4^*=120.$$

The corresponding value of the objective function which is the value of the weighted sum of $V(\bar{y}_{jst})$ (f.p.c. ignored) is $Z^*=69.0801$.

5.5 DISCUSSION

In the numerical illustration presented in the section 5.4 the total sample size $n=\sum_{h=1}^{4} n_h=515$. As suggested by Neyman (1934), if proportional allocation is used, with n=515 and values of W_h given in Table 5.1 we get the sample size $n_h=nW_h$; h=1,2,3 and 4 as:

$$n_1=174,\ n_2=76,\ n_3=154 \text{ and } n_4=111.$$

Table 5.2 gives the values of $V(\bar{y}_{jst}) = \dfrac{\sum_{h=1}^{4} W_h S_{jh}^2}{n}$, j=1&2 under the proportional allocation (ignoring fpc).

Table 5.2

Variance of $\bar{y}_{jst}$ under proportional allocation, ignoring fpc, for total sample size n=515

h	W_h	j=1 $W_h S_{h1}^2$	j=2 $W_h S_{h2}^2$
1	0.3387	1631.7618	44072.0335
2	0.1477	923.3111	1124.5169
3	0.2990	916.7818	435.4636
4	0.2146	12062.0758	14373.4187
	$\sum$	15533.9305	60005.4327
	$V(\bar{y}_{jst})$	30.1630	116.5154

Under the proportional allocation the weighted sum of variances is worked out as:

$$\sum_{j=1}^{2} a_j V(\bar{y}_{jst}) = 0.25 \times 30.1630 + 0.75 \times 116.5154$$

$$= 94.9273$$

The relative efficiency (R.E.) of the integer compromise allocation as compared to the proportional allocation is

$$R.E. = \frac{\sum_{j=1}^{p} a_j V(\bar{y}_{jst})_{prop}}{\sum_{j=1}^{p} a_j V(\bar{y}_{jst})_{comp}} \times 100\%$$

$$= \frac{94.9273}{69.0801} \times 100\%$$

$$= 137.42\%$$

Which shows that the proposed procedure provides an allocation which is more precise than the usual proportional allocation.

REFERENCES

Aggarwal, O.P. (1974a), "On mixed integer quadratic problems", *Naval Research Logistics Quarterly*, **21**.

Aggarwal, O.P. (1974b), "On integer solution to quadratic problems by a branch and bound technique", *Trabajas de Estatistica Y De Investigation Operation*, **25**, 65-70.

Ahsan, M.J. and Khan, S.U. (1977), "Optimum allocation in multivariate stratified random sampling using prior information", *J. Ind. Stat. Assoc.*, **15**, 57-67.

Ahsan, M.J. and Khan, S.U. (1982), "Optimum allocation in multivariate stratified random sampling with overhead cost", *Metrika*, **29**, 71-78.

Ahsan, M.J.; Khan, S.U. and Arshad, M. (1983), "Minimizing a non-linear function arising in stratification through approximation by a quadratic function", *J. Ind. Soc. Statist. Oper. Res.*, **4**, 1-4, 9-16.

Anstreicher, K.M.; den Hertog, D. and Terlaky, T. (1994), "A long-step barrier method for convex quadratic programming", *Algorithmica*, **10**, 5, 365-382.

Aoyama, H. (1954), "A study of stratified random sampling", *Ann. Inst. Stat. Math.*, **6**, 1-36.

Aoyama, H. (1963), "Stratified random sampling with optimum allocation for multivariate populations", *Ann. Inst. Stat. Math.*, **14**, 251-258.

Arshad, M.; Khan, S.U. and Ahsan, M.J. (1981), "A procedure for solving concave quadratic programs", *Aligarh Journal of Statistics*, **1**, 2, 106-112.

Arthanari, T.S. and Dodge, Y. (1981), "*Mathematical Programming in Statitistics*", Wiley, New York.

Beal, E.M.L. (1959), "On quadratic programming", *Naval Research Logistics Quarterly*, **6**, 227-243.

Bellman, R.E. (1957), "*Dynamic Programming*", Princeton University Press, Princeton.

Bellman, R.E. and Dreyfus, S.E. (1962), "*Applied Dynamic Programming*", Princeton University Press, Princeton.

Ben-Daya, M. and Shetty, C.M. (1990), "Polynomial barrier function algorithms for convex quadratic programming", *Arabian J. Sci. Engrg.*, **15**, No. 4, B, 657-670.

Benzi, M. (1993), "Solution of equality-constrained quadratic programming problems by a projective iterative method", *Rend. Mat. Appl.* (7), **13**, 2, 275-296.

Bomze, I.M. and Danninger, G. (1993), "A global optimization algorithm for concave quadratic programming problems", *SIAM J. Optim.*, **3**, 4, 826-842.

Bomze, I.M. and Danninger, G. (1994), "A finite algorithm for solving general quadratic problems", *J. Global Optim.*, **4**, 1, 1-16.

Chaddha, R.L.; Hardgrave, W.W.; Hudson, D.J.; Segal, M. and Suurballe, J.W. (1971), "Allocation of total sample size when only the stratum means are of interest", *Technometrices*, **13**, 4, 817-831.

Chatterjee, S. (1967), "A note on optimum allocation", *Skand. Akt.*, **50**, 40-44.

Chatterji, S. (1968), Multivariate stratified surveys", *J. Amer. Stat. Assoc.*, **63**, 530-534.

Chen, H.-D.; Hearn, D.W. and Lee, C.-Y. (1994), "A dynamic programming algorithm for dynamic lot size models with piecewise linear cost", *J. Global Optim.*, **4**, No. 4, 397-413.

Cochran, W.G. (1961), "Comparision of methods for determining stratum boundaries", *Bull. Int. Stat. Inst.*, **38**, 2, 345-358.

Cochran, W.G. (1963), "*Sampling Techniques*", (2nd ed), Wiley, New York.

Cochran, W.G. (1977), "*Sampling techniques*" (3rd ed) John Wiley and Sons, Inc., New York.

Dalenius, T. (1950), "The problem of optimum stratification-I", *Skand. Akt.*, **33**, 203-213.

Dalenius, T. and Gurney, M. (1951), "The problem of optimum stratification-II", *Skand. Akt.*, **34**, 133-148.

Dalenius, T. (1953), "Multivariate sampling problem", *Skand. Akt.*, **36**, 92-122.

Dalenius, T. (1957), "*Sampling in Sweden*: Contributions to the Methods and Theories of Sample Survey Practice", Almqvist Och Wiksell, Stockholm.

Dalenius, T. and Hodges, J.L. (1959), "Minimum variance stratification", *Jour. Amer. Stat. Assoc.*, **54**, 80-101.

Du, D.-Z.; Wu, F. and Zhang, X.-S. (1990), "On Rosen's gradient projection methods", *Anal. Oper. Res.*, **24**, 1-4, 11-28.

Durbin, J. (1959), "Review of sampling in Sweden", *Jour. Roy. Stat. Soc.*, A, **122**, 246-248.

Finkbeiner, B. and Kall, P. (1978), "Direct algorithm in quadratic programming", *Zeitschrift fur Operation Research*, **17**, 45-54.

Fletcher, R. (1971), "A general quadratic programming algorithm", *Jour. of the Inst. of Mathematics and its Applications*, **7**, 76-91.

Fletcher, R. (1993), "Resolving degeneracy in quadratic programming", *Ann. Oper. Res.*, **46/47**, No. 1-4, 307-334.

Folks, John Leroy and Antle, Charles E. (1965), "Optimum allocation of sampling units to the strata when there are R responses of interest", *J. Amer. Statist. Assoc.*, **60**, 225-233.

Fukushima, M. (1986), "A successive quadratic programming algorithm with global and superlinear convergence properties", *Mathematical Programming*, **35**, No. 3, 253-264.

Geary, R.C. (1949), "Sampling methods applied to Irish agricultural statistics", *Technical Series*, Sept. 1949.

Ghosh, S.P. (1958), "A note on stratified random sampling with multiple characters", *Calcutta Stat. Assoc. Bull.*, **8**, 81-89.

Goldfarb, D. (1969), "Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints", *SIAM J. Appl. Math.*, **17**, 739-764.

Graves, R.L. (1967), "A principal pivoting simplex algorithm for linear and quadratic programming", *Operation Research*, **15**, 482-494.

Hadley, G. (1964), "*Nonlinear and Dynamic Programming*", Addison-Wesley Publishing Company, Inc., London.

Hansen, M.H.; Hurwitz, W.N. and Madow, W.G. (1953), "*Sample Survey Methods and Theory*", Wiley, New York, Vol. 1.

Hess, J.; Sethi, V.K. and Balakrishnan, T.R. (1966), "Stratification: A practical investigation", *J. Amer., Stat. Assoc.*, **61**, 74-90.

Jahan, N.; Khan, M.G.M. and Ahsan, M.J. (1994), "A generalized compromise allocation", *J. Indian Statistical Association*, Vol. 32, No. 2, 95-101.

Jessen, R.J. (1942), "Statistical investigation of a sample survey for obtaining farm facts", *Iowa Agr. Exp. Stat. Res. Bull.*, **304**.

Kalantari, B. and Bagchi, A. (1990), "An algorithm for quadratic zero-one programs", *Naval Research Logistics*, **37**, No. 4, 527-538.

Kelley, J.E. (1960), "The cutting-plane method for solving convex programs", *J. Soc. Indust. Appl. Math.*, **8**, 703-712.

Khan, E.A.; Khan, M.G.M. and Ahsan, M.J. (1997), "An optimal multivariate stratified sampling design using dynamic programming presentedd in the "3rd International Triennial Calcutta Symposium" held in December 1997, Proceedings of the symposium to be published by Wiely Publications (to appear).

Khan, E.A.; Khan, M.G.M. and Ahsan, M.J. (1998), "Optimum stratification: a mathematical programming approach", accepted for presentation at the sixth Islamic Countries Conference on Statistical Sciences to be held in Dhaka, Bangladesh during December 12-15, 1998.

Khan, M.G.M.; Ahsan, M.J. and Khan, E.A. (1997), "On compromise allocation in multivariate stratified sampling", submitted for publication to Naval Research Logistics (vide their manuscript number 3280).

Khan, M.G.M.; Ahsan, M.J. and Jahan, N. (1997), "Compromise allocation in multivariate stratified sampling: an integer solution", *Naval Research Logistics*, **44**, 69-79.

Khan, M.G.M.; Khan, E.A. and Jahan, N. (1998), "Determining the optimum number of strata II", *Frontiers in Probability and Statistics*, Edited by S.P. Mukherjee, S.K. Basu, and B.K. Sinha, Narosa Publishing House, 215-225.

Khan, Z.A.; Ahsan, M.J. and Khan, E.A. (1983), "On mixed integer concave quadratic programs", *Soochow Journal of Mathematics*, 9, 111-116.

Kokan, A.R. and Khan, S.U. (1967), "Optimum allocation in multivariate surveys: an analytical solution", *J. Roy. Statist. Soc.*, Ser. B, **29**, 115-125.

Kuhn, H.W. and Tucker, A.W. (1951), "Nonlinear programming, in proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 481-492.

Lai, Y.L.; Gao, Z.Y. and He, G.P. (1993), "A generalized gradient projection algorithm of optimization with nonlinear constraints", *Sci. China.*, Ser. A, **36**, No. 2, 170-180.

Lemke, C.E. (1962), "A method for solution of quadratic programs", *Manage Science*, **8**, 442-453.

Li, D. (1990), "Multiple objective and nonseparability in stochastic dynamic programming", *Internat. J. Systems Sci.*, **21**, No. 5, 933-950.

Li, D. and Haimes, Y.Y. (1990), "New approach for nonseparable dynamic programming problems", *J. Optim. Theory Appl.*, **64**, No. 2, 311-330.

Mahalanobis, P.C. (1952), "some aspects of the design of sample surveys", *Sankhya*, **12**, 1-17.

Murthy, M.N. (1967), "*Sampling Theory and Methods*", Statistical Publishing Society, Calcutta.

Neyman, J. (1934), "On the two different aspects of the representative methods: the method of stratified sampling and the method of purposive selection", *J. Roy. Stat. Soc.*, **97**, 558-606.

Odanaka, T. (1994), "Dynamic programming and optimal inventory processes", *Comput. Math. Appl.*, **27**, No. 9-10, 213-217.

Omule, S.A.Y. (1958), "Optimum design in multivariate stratified sampling", *Biom. J.*, **27**, 8, 907-912.

Peter, J.H. and Bucher, (Undated), "*The 1940 Section Sample Survey of Crop Aggregates in Indiana and Iowa*", U.S., Dept. of Agriculture.

Powell, M.J.D. and Yuan, Y. (1986), "A recursive quadratic programming algorithm that uses differentiable exact penalty functions", *Mathematical Programming*, **35**, No. 3, 265-278.

Rosen, J.B. (1960), "The gradient projection method for nonlinear programming, Part I: linear constraints", *J. Soc. Indust. Appl. Math.*, **8**, 181-217.

Rosen, J.B. (1961), "The gradient projection method for nonlinear programming, Part II: nonlinear constraints", *J. Soc. Indust. Appl. Math.*, **9**, 514-532.

Sethi, V.K. (1963), "A note on optimum stratification of population for estimating the population mean", *Aust. Jour. Stat.*, **5**, 20-23.

Sukhatme, P.V.; Sukhatme, B.V.; Sukhatme, S. and Asok, C. (1984), "*Sampling Theory of Surveys with Applications*", Iowa State University Press, Ames, Iowa (U.S.A) and Ind. Soc. of Agr. Stats., New Delhi (India).

Todd, M.J. (1985), "Linear and quadratic programming in oriented matroids", *J. of Combinatorial Theory*, B. **39**, 105-133.

Unnithan, V.K.G. (1978), "The minimum variance boundary points of stratification", *Sankhya*, **40**, C, 60-72.

Van de Panne, C. and Whinston, A. (1964a), "Simplical methods for quadratic programming", *Naval Research Logistics Quarterly*, **11**, 273-302.

Van de Panne, C. and Whinston, A. (1964b), "Simplex and dual methods for quadratic programming", *Operations Research Quarterly*, **15**, 355-388.

Van de Panne, C. and Whinston, A. (1966), "The symmetric function of the simplex method for quadratic programming", *Paper*, **112**, Western Management Science Institute, University of California, Los Angeles.

Wachs, M.L. (1989), "On an efficient dynamic programming technique of F.F. Yao", *J. Algorithms*, **10**, No. 4, 518-530.

Wang, C.-L. (1990a), "Dynamic programming and inequality", *J. Math. Anal. Appl.*, **150**, No. 2, 528-555.

Wang, C.-L. (1990b), "Dynamic programming and the Lagrange multipliers", *J. Math. Anal. Appl.*, **150**, No. 2, 551-561.

Wang, C.-L. and Xing, A.-Q. (1990), "Dynamic programming and penalty functions", *J. Math. Anal. Appl.*, **150**, No. 2, 562-573.

Wei, Z.X. (1992), "A new successive quadratic programming algorithm", *Acta Math. Appl. Sinica* (English Ser.), **8**, No. 3, 281-288.

Wolfe, P. (1959), "The simplex method for quadratic programming", *Econometrica*, **27**, 382-398.

Yates, F. (1960), "*sampling Methods for Censuses and Surveys*", (2nd ed), Charles Griffin and Co. Ltd., London.

Yuan, Y.X. (1991), "A dual algorithm for minimizing a quadratic function with two quadratic constraints", *J. Comput. Math.*, **9**, No. 4, 348-359.